Degenerate elliptic resonances

Guido Gentile† **and Giovanni Gallavotti***

† Dipartimento di Matematica, Università di Roma Tre, Roma, I-00146

* Dipartimento di Fisica, Università di Roma "La Sapienza", I-00185

Abstract. *Quasi-periodic motions on invariant tori of an integrable system of dimension smaller than half the phase space dimension may continue to exists after small perturbations. The parametric equations of the invariant tori can often be computed as formal power series in the perturbation parameter and can be given a meaning via resummations. Here we prove that, for a class of elliptic tori, a resummation algorithm can be devised and proved to be convergent, thus extending to such lower-dimensional invariant tori the methods employed to prove convergence of the Lindstedt series either for the maximal (i.e. KAM) tori or for the hyperbolic lower-dimensional invariant tori.*

1. Introduction

Quasi-integrable analytic Hamiltonian systems are described by Hamiltonians of the form $\mathcal{H} = \mathcal{H}_0(\mathbf{I}) + \varepsilon \mathcal{H}_1(\boldsymbol{\varphi}, \mathbf{I})$, where $(\boldsymbol{\varphi}, \mathbf{I}) \in \mathbb{T}^d \times \mathcal{A}$, with $\mathcal{A}$ an open domain in $\mathbb{R}^d$, are conjugate coordinates (called angle-action variables), the functions $\mathcal{H}_0$ and $\mathcal{H}_1$ are analytic in their arguments, and ε is a small real parameter. We shall consider for simplicity only Hamiltonians of the form

$$\mathcal{H} = \frac{1}{2}\mathbf{I} \cdot \mathbf{I} + \varepsilon f(\boldsymbol{\varphi}), \tag{1.1}$$

where $\cdot$ denotes the inner product in $\mathbb{R}^d$.

Kolmogorov's theorem (KAM theorem) yields, for ε small enough, the existence of many invariant tori for Hamiltonian systems of the form (1.1): such tori can be parameterized by the corresponding rotation vectors, at least if the latter satisfy some Diophantine conditions. On the other hand Poincaré's theorem states the existence of periodic orbits, which can be parameterized by rotation vectors satisfying $d-1$ resonance conditions (so that after a simple linear canonical map one can assume that the rotation vector is $(\omega_1, 0, 0, \ldots, 0)$).

A natural question is what happens of the invariant tori corresponding, in absence of perturbations, to rotation vectors satisfying s resonance conditions, with $1 \le s \le d-2$. If we fix the rotation vector as $(\boldsymbol{\omega}, \mathbf{0}) \equiv (\omega_1, \ldots, \omega_r, 0, \ldots, 0)$ and parameterize the invariant torus for $\varepsilon = 0$ with the action value $\mathbf{I} = \mathbf{0}$ then, after translating the origin in $\mathbb{R}^d$ by $(\boldsymbol{\omega}, \mathbf{0})$ and setting $\mathbf{I} = (\mathbf{A}, \mathbf{B}) \in \mathbb{R}^r \times \mathbb{R}^s$, $\boldsymbol{\varphi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{T}^r \times \mathbb{T}^s$, the Hamiltonian (1.1) becomes

$$\mathcal{H} = \boldsymbol{\omega} \cdot \mathbf{A} + \frac{1}{2}\mathbf{A} \cdot \mathbf{A} + \frac{1}{2}\mathbf{B} \cdot \mathbf{B} + \varepsilon f(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{1.2}$$

where $(\boldsymbol{\alpha}, \mathbf{A}) \in \mathbb{T}^r \times \mathbb{R}^r$ and $(\boldsymbol{\beta}, \mathbf{B}) \in \mathbb{T}^s \times \mathbb{R}^s$ are conjugate variables, with $r+s=d$, and $\cdot$ denotes

the inner product both in $\mathbb{R}^r$ and in $\mathbb{R}^s$. Here we impose that $\boldsymbol{\omega}$ is a vector in $\mathbb{R}^r$ satisfying

$$|\boldsymbol{\omega}\cdot\boldsymbol{\nu}| \ge C_0|\boldsymbol{\nu}|^{-\tau_0} \qquad \forall\,\boldsymbol{\nu}\in\mathbb{Z}^r\setminus\{\mathbf{0}\}, \tag{1.3}$$

with $C_0>0$ and $\tau_0\ge r-1$, which is called the *Diophantine condition*; we shall define by $D_{\tau_0}(C_0)$ the set of rotation vectors in $\mathbb{R}^r$ satisfying (1.3). We also write

$$f(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{\boldsymbol{\nu}\in\mathbb{Z}^r} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\alpha}} f_{\boldsymbol{\nu}}(\boldsymbol{\beta}). \tag{1.4}$$

We shall suppose that f is analytic in a strip around the real axis of the variables $\boldsymbol{\alpha},\boldsymbol{\beta}$, so that there exist constants F_0, F_1, κ_0 such that $|\partial_{\boldsymbol{\beta}}^q f_{\boldsymbol{\nu}}(\boldsymbol{\beta})| \le q! F_0 F_1^q e^{-\kappa_0|\boldsymbol{\nu}|}$ for all $\boldsymbol{\nu}\in\mathbb{Z}^r$ and all $\boldsymbol{\beta}\in\mathbb{T}^s$.

There are quite a few results on the above problem, essentially solved, under the assumptions of Theorem 1 below, in Ref. [JLZ], and on closely related problems. We summarize our understanding of the existing results in Appendix A1.

The equations of motion for the system (1.2), written in terms of the angle variables alone, are

$$\ddot{\boldsymbol{\alpha}} = -\varepsilon\partial_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha},\boldsymbol{\beta}), \qquad \ddot{\boldsymbol{\beta}} = -\varepsilon\partial_{\boldsymbol{\beta}} f(\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{1.5}$$

so that, once a solution of (1.5) is found, the action variables are immediately obtained by a simple differentiation: $\mathbf{A}=\dot{\boldsymbol{\alpha}}-\boldsymbol{\omega}$, $\mathbf{B}=\dot{\boldsymbol{\beta}}$.

We look for solutions of (1.5), for $\varepsilon\neq 0$, conjugated to the free solution $(\boldsymbol{\alpha}_0+\omega t,\boldsymbol{\beta}_0,0,0)$, i.e. we look for solutions of the form

$$\boldsymbol{\alpha}(t) = \boldsymbol{\psi} + \mathbf{a}(\boldsymbol{\psi},\boldsymbol{\beta}_0;\varepsilon), \qquad \boldsymbol{\beta}(t) = \boldsymbol{\beta}_0 + \mathbf{b}(\boldsymbol{\psi},\boldsymbol{\beta}_0;\varepsilon), \tag{1.6}$$

for some functions $\mathbf{a}$ and $\mathbf{b}$, real analytic and 2π-periodic in $\boldsymbol{\psi}\in\mathbb{T}^r$, such that the motion in the variable $\boldsymbol{\psi}$ is governed by the equation $\dot{\boldsymbol{\psi}}=\boldsymbol{\omega}$. We shall prove the following result.

Theorem 1. *Consider the Hamiltonian (1.2), with $\boldsymbol{\omega}\in D_{\tau_0}(C_0)$ and f analytic and periodic in both variables. Suppose $\boldsymbol{\beta}_0$ to be such that*

$$\partial_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0) = \mathbf{0}, \tag{1.7}$$

and assume that the eigenvalues $a_1,\ldots,a_s$ of the matrix $\partial_{\boldsymbol{\beta}}^2 f_{\mathbf{0}}(\boldsymbol{\beta}_0)$ are pairwise distinct and positive, i.e. for some constant $a>0$ one has a_i, $a_j-a_i > a > 0$ for all $j>i=1,\ldots,s$.
Then there exist a constant $\overline{\varepsilon}>0$ and a set $\mathcal{E}\subset(0,\overline{\varepsilon})$ such that the following holds.
(i) For all $\varepsilon\in\mathcal{E}$ there are solutions of (1.5) of the form (1.6), where the two functions $\mathbf{a}(\boldsymbol{\psi},\boldsymbol{\beta}_0;\varepsilon)$ and $\mathbf{b}(\boldsymbol{\psi},\boldsymbol{\beta}_0;\varepsilon)$ are real analytic and 2π-periodic in the variables $\boldsymbol{\psi}\in\mathbb{T}^r$.
(ii) The relative Lebesgue measure of $\mathcal{E}\cap(0,\varepsilon)$ with respect to $(0,\varepsilon)$ tends to 1 for $\varepsilon\to 0$.
(iii) The functions $\mathbf{a},\mathbf{b}$ can be extended to Lipschitz functions of $\varepsilon,\boldsymbol{\psi}$ in $[0,\overline{\varepsilon}]\times\mathbb{T}^r$.

Remarks. (1) From the literature one might expect that the non-resonance condition on the eigenvalues of $\partial_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0)$ could be avoided; see Appendix A1.
(2) The case of negative ε was dealt with in Ref. [GG], with techniques close to the ones introduced here, and it corresponds to the case of hyperbolic tori.
(3) The case of mixed stationarity, i.e. $\det\partial_{\boldsymbol{\beta}}^2 f_{\mathbf{0}}(\boldsymbol{\beta}_0)\neq 0$ and eigenvalues of $\partial_{\boldsymbol{\beta}}^2 f_{\mathbf{0}}(\boldsymbol{\beta}_0)$ of mixed signs (with non-degeneracy of the positive ones), can be treated in exactly the same way discussed

in this paper and the above result extends to this case; cf. Theorem 2 in Section 7.
(4) For $\varepsilon \notin \mathcal{E}$ the smooth extension in (iii) does not represent parametric equations of invariant tori: it just says that their values in the physically interesting set $\mathcal{E}$ (which turns out to have dense complement in $[0,\overline{\varepsilon}]$) can be smoothly interpolated in ε. Such (non-unique) extensions are commonly used for interpolation purposes and are called Whitney extensions.

The novelty and the purpose of our work is the development of a method of proof based on the existence of a formal power series expression for the functions $(\mathbf{a},\mathbf{b})$ *and its multiscale analysis producing a rearrangement of its terms, involving summing many divergent series, which turns it into an absolutely convergent series.*

The paper is organized as follows. In Section 2 we recall the basic formalism, following Ref. [GG], and in Section 3 we give a simple example of resummation.

In Section 4 we set up terminology and discuss *heuristically* the ideas governing our resummations, by explaining why they have to be performed by a multiscale analysis of the series (which we call "Lindstedt series") representing a formal expansion of the quasi-periodic motions in powers of ε.

The singularities are first "probed" down to a scale in which possible resonances between the proper frequencies, i.e. the components of $\boldsymbol{\omega}$, and the normal frequencies, i.e. the square roots of the eigenvalues of $\varepsilon\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0)$, are still irrelevant. The analysis of such singularities leads to what we call *non-resonant* or *high frequency resummations*, which can be treated by the method of Ref. [GG], i.e. of the hyperbolic case, in which no resonances at all were possible between proper frequencies and normal frequencies (simply because, for the Hamiltonian (1.2) the latter did not exist). Further probing of the singularities leads to what we call the *resonant* (or *infrared*) *resummations*: the analysis is more elaborated and it requires new ideas, obtained by combining the ideas in Ref. [GG] with the ones introduced in Ref. [Ge].

In Section 5 we discuss the non-resonant resummations while the new infrared resummations are studied in Section 6 where a "fully renormalized series" is obtained, i.e. a resummation of the series defining the formal expansion of the quasi-periodic solution (1.6) of the equations of motion (1.5), *which we prove to be absolutely convergent.*

The paper is a self-contained discussion of the construction and of the convergence of the resummed series. This includes a self-contained *description* of the well-known formal series, [JLZ], [GG]. Once this is achieved one has to check that the defined functions do actually represent parametric equations of invariant tori: for this we follow, in Appendix A5, the analysis of Refs. [GG] and [Ge].

2. Tree formalism

We look for a formal power series expansion (in ε) of the parametric equations $\mathbf{h} = (\mathbf{a},\mathbf{b})$ of the invariant torus close to the torus $\boldsymbol{\alpha} = \boldsymbol{\psi}, \boldsymbol{\beta} = \boldsymbol{\beta}_0$

$$\mathbf{h}(\boldsymbol{\psi};\varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k \mathbf{h}^{(k)}(\boldsymbol{\psi}) = \sum_{\boldsymbol{\nu}\in\mathbb{Z}^r} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}} \mathbf{h}_{\boldsymbol{\nu}}(\varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu}\in\mathbb{Z}^r} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}}\, \mathbf{h}^{(k)}_{\boldsymbol{\nu}}, \tag{2.1}$$

where we have not explicitly written the dependence on $\boldsymbol{\beta}_0$. The power series is easy to derive, see for instance Ref. [GG]: however its convergence turns out to be substantially harder than the convergence proof of the Lindstedt series for the maximal KAM tori. The series constructed below

for our problem, which we still call "Lindstedt series", is naturally described in terms of trees. The coefficients $\mathbf{h}_{\boldsymbol{\nu}}^{(k)}$ can be computed as sums of "values" that we attribute to trees whose nodes and lines carry a few labels, which we call "decorated trees".

The formalism to define trees, decorations and values has been described many times and used in the proof of several stability results in Hamiltonian mechanics. Usage of graphical tools based on trees in the context of KAM theory has been advocated recently in the literature as an interpretation of Ref. [E1]; see for instance Refs. [Ga], [GG], [BGGM] and [BaG].

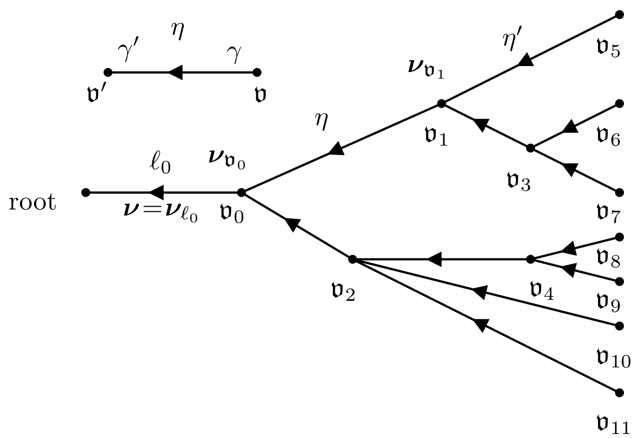

FIGURE 1. A tree θ with 12 nodes; one has $p_{\mathfrak{v}_0} = 2, p_{\mathfrak{v}_1} = 2, p_{\mathfrak{v}_2} = 3, p_{\mathfrak{v}_3} = 2, p_{\mathfrak{v}_4} = 2$. The length of the lines should be the same but it is drawn of arbitrary size. The endnodes $\mathfrak{v}_i$, $i = 5, \ldots, 11$ can be either nodes or leaves of the tree. The separated line illustrates the way to think of the label $\eta = (\gamma', \gamma)$.

A tree θ (see Fig. 1) is defined as a partially ordered set of points, connected by oriented *lines*. The lines are consistently oriented toward the *root*, which is the leftmost point $\mathfrak{r}$; the line entering the root is called the *root line*. If a line ℓ connects two points $\mathfrak{v}_1, \mathfrak{v}_2$ and is oriented from $\mathfrak{v}_2$ to $\mathfrak{v}_1$, we say that $\mathfrak{v}_2 \prec \mathfrak{v}_1$ and we shall write $\mathfrak{v}_2' \stackrel{def}{=} \mathfrak{v}_1$ and $\ell_{\mathfrak{v}_2} \stackrel{def}{=} \ell$; we shall say also that ℓ exits $\mathfrak{v}_2$ and enters $\mathfrak{v}_1$. More generally we write $\mathfrak{v}_2 \prec \mathfrak{v}_1$ if $\mathfrak{v}_2$ is on the path of lines connecting $\mathfrak{v}_1$ to the root. The points different from the root will be called the *nodes* of the tree.

Each line from $\mathfrak{v}$ to $\mathfrak{v}'$ carries a pair η of labels $\eta = (\gamma, \gamma')$ ranging in $\{1, \ldots, d\}$ (marked in Fig. 1 only on some of the lines for clarity of the drawing). The labels γ and γ' should be regarded as associated with $\mathfrak{v}$ and $\mathfrak{v}'$, respectively; hence with each node $\mathfrak{v}$ with $p_{\mathfrak{v}}$ entering lines $\ell_1, \ldots, \ell_{p_{\mathfrak{v}}}$ one can associate $p_{\mathfrak{v}} + 1$ labels $\gamma_0, \gamma_1, \ldots, \gamma_{p_{\mathfrak{v}}}$, with $\gamma_0 = \gamma_{\ell_{\mathfrak{v}}}$ and $\gamma_j = \gamma'_{\ell_i}$. Also the root line (from $\mathfrak{v}_0$ to the root) carries two such labels and the one associated with the final extreme of the root line will be called the *root label*.

Fixed any $\ell_{\mathfrak{v}} \in \theta$, we shall say that the subset of θ containing $\ell_{\mathfrak{v}}$ as well as all nodes $\mathfrak{w} \preceq \mathfrak{v}$ and all lines connecting them is a *subtree* of θ with root $\mathfrak{v}'$; of course a subtree is a tree.

Given a tree, with each node $\mathfrak{v}$ we associate a *harmonic* or *mode*, as called in Ref. [GG], i.e. a label $\boldsymbol{\nu}_{\mathfrak{v}} \in \mathbb{Z}^r$. We shall denote by $V(\theta)$ the set of nodes and by $\Lambda(\theta)$ the set of lines. The number $k = |V(\theta)|$ of nodes in the tree θ, equal to the number $|\Lambda(\theta)|$ of lines, will be called the *order* of θ.

We call a node with one entering line and $\mathbf{0}$ harmonic label a *trivial node*.

With any line $\ell = \ell_{\mathfrak{v}}$ we associate (besides the above mentioned pair $\eta_\ell = (\gamma_\ell, \gamma'_\ell)$ of labels

assuming values in $\{1,\ldots,d\}$) a *momentum label* $\boldsymbol{\nu}_\ell \in \mathbb{Z}^r$ defined as

$$\boldsymbol{\nu}_\ell \equiv \boldsymbol{\nu}_{\ell_{\mathfrak{v}}} = \sum_{\substack{\mathfrak{w}\in V(\theta)\\ \mathfrak{w}\preceq\mathfrak{v}}} \boldsymbol{\nu}_{\mathfrak{w}}. \tag{2.2}$$

We shall assume that no tree contains trivial nodes with the entering line with $\mathbf{0}$ *momentum*: this is an important restriction, as we shall see. We call *degree* $P(\theta)$ of a tree the order of the tree minus the number of $\mathbf{0}$ momentum lines, so that $|V(\theta)| - P(\theta)$ is their number.

We call $\Theta_{\boldsymbol{\nu},k,\gamma}$ the set of trees θ of order k, i.e. with $|V(\theta)| = k$ nodes, and $\Theta^o_{\boldsymbol{\nu},k,\gamma}$ the set of trees of degree k, i.e. with $P(\theta) = k$. One has $\Theta_{\boldsymbol{\nu},k,\gamma} \neq \Theta^o_{\boldsymbol{\nu},k,\gamma}$.

Each tree θ "decorated" by labels in the described way will have a *value* which is defined in terms of a product of several factors.

• With each node $\mathfrak{v}$ we associate a *node factor*

$$F_{\mathfrak{v}} = \prod_j \partial_{\gamma_j} f_{\boldsymbol{\nu}_{\mathfrak{v}}}(\boldsymbol{\beta}_0), \tag{2.3}$$

where the labels γ_j are the $p_{\mathfrak{v}}+1$ labels associated with the extreme $\mathfrak{v}$ of the $p_{\mathfrak{v}}$ lines entering the node $\mathfrak{v}$ and of the line exiting it, and the derivatives ∂_γ, with $\gamma = 1,2,\ldots,r$, have to be interpreted as factors $(i\boldsymbol{\nu}_{\mathfrak{v}})_\gamma$. Hence $F_{\mathfrak{v}}$ is a tensor of rank $p_{\mathfrak{v}}+1$.

• With each line ℓ carrying labels $\eta_\ell = (\gamma_\ell, \gamma_{\ell'})$ and momentum $\boldsymbol{\nu}_\ell$ we associate a matrix, called *propagator*,

$$\begin{aligned} G_\ell &\equiv \delta_{\gamma_\ell,\gamma'_\ell}\frac{1}{(\boldsymbol{\omega}\cdot\boldsymbol{\nu}_\ell)^2}, && \text{if } \boldsymbol{\nu}_\ell \neq \mathbf{0},\\ G_\ell &\equiv -\varepsilon^{-1}\,(\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0))^{-1}_{\gamma_\ell,\gamma'_\ell}\,\chi(\gamma_\ell,\gamma'_\ell > r), && \text{if } \boldsymbol{\nu}_\ell = \mathbf{0}, \end{aligned} \tag{2.4}$$

where $\chi(\gamma_\ell,\gamma'_\ell > r)$ is 1 if both γ_ℓ and γ'_ℓ are strictly greater than r, and 0 otherwise.

Given the definitions (2.3) and (2.4) define a *value function* Val, which with each tree θ of order k associates a *tree value*

$$\mathrm{Val}(\theta) = \frac{\varepsilon^k}{k!}\Big(\prod_{\mathfrak{v}\in V(\theta)} F_v\Big)\Big(\prod_{\ell\in\Lambda(\theta)} G_\ell\Big), \tag{2.5}$$

where, by the definitions, all labels γ_i associated with the nodes appear twice because they appear also in the propagators: we make in (2.5) the *summation convention* that repeated γ labels associated with nodes and lines are summed over, with the exception of the label γ associated with the root (because we do not consider it a node and the corresponding label γ appears only once in (2.5)). Therefore (2.5) is a number labeled by $\gamma = 1,\ldots,d$, i.e. $\mathrm{Val}(\theta)$ is a vector.

Remarks. (1) The trees can be drawn in various ways: we can limit the arbitrariness by demanding that the length of the segments representing the lines is 1 (unlike the drawings in the above figures) and that the angles between the lines are irrelevant. *The combinatorics being very important*, because it matters in the check of cancellations essential for the analysis, we adopt the convention that trees are drawn on a plane, have lines of unit length *and carry an identifier label*, that we call *number label* (not shown in the above figures) which distinguishes the lines from each other even if we ignore the other labels attached to them. Furthermore two trees that can be superposed by pivoting the lines merging into the same node $\mathfrak{v}$, around $\mathfrak{v}$ itself, are considered *identical.* This is a convention which is useful for checking cancellations: however it is by no means the only possible

one. Others are possible and often very convenient in other respects [G3], [GM1], but in a given work a choice has to be made once and for all.
(2) A line ℓ carrying $\mathbf{0}$ momentum is somewhat special. We could visualize the part of the tree preceding such lines by encircling it into a dotted circle: such a representation has been used in earlier papers, e.g. in Ref. [GG], calling the subtree θ_ℓ with ℓ as root line a *leaf*. Here, however, we shall avoid using a special word for the $\mathbf{0}$ momentum lines and the subtrees preceding them.
(3) We can think of the propagators as matrices of the form

$$G_\ell = \begin{pmatrix} G_{\ell,\alpha\alpha} & G_{\ell,\alpha\beta} \\ G_{\ell,\beta\alpha} & G_{\ell,\beta\beta} \end{pmatrix}, \tag{2.6}$$

where $G_{\ell,\alpha\alpha}$, $G_{\ell,\alpha\beta}$, $G_{\ell,\beta\alpha}$ and $G_{\ell,\beta\beta}$ are $r\times r$, $r\times s$, $s\times r$ and $s\times s$ matrices.
(4) The value of a tree θ defined above has no pole at $\varepsilon=0$ if $\mathrm{Val}(\theta)\neq 0$ because every line with $\mathbf{0}$ momentum is preceded by at least two nodes, so that the total power of ε to which the value is proportional is always non-negative and, in fact, it is necessarily positive: we need to take into account that $\partial_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0)\equiv 0$ and that our trees contain no trivial nodes with one entering line with $\mathbf{0}$ momentum. Note that $\mathrm{Val}(\theta)$ is a monomial in ε of degree $P(\theta)$.
(5) In the case of maximal tori and if $\mathrm{Val}(\theta)\neq 0$ there are *no lines with* $\mathbf{0}$ *momentum* for systems described by the Hamiltonians (1.1): indeed $s=0$, see also [Ga]. In this case the number of nodes, i.e. the tree order, coincides with the power of ε associated with the monomial in ε defined by the tree value, i.e. with the tree degree. In general, however, the order $|V(\theta)|$ of a tree can be larger than its degree $P(\theta)$: $|V(\theta)|\geq P(\theta)\geq \frac12|V(\theta)|$.

The above definitions uniquely attribute a value to each tree. The following result states the existence of formal solutions to (1.5) which are conjugated to the unperturbed motion (1.4), provided the value $\boldsymbol{\beta}_0$ is suitably fixed. The proof is an algebraic check which does not distinguish the possible signs of ε and can be taken from Ref. [GG] where it is done in the case $\varepsilon<0$.

LEMMA 1. *The Fourier transform of the power series solution* $\mathbf{h}=(\mathbf{a},\mathbf{b})$ *of (1.5) of the form (2.1) is obtained by writing (the definition of* $\Theta^o_{k,\boldsymbol{\nu},\gamma}$ *follows (2.2))*

$$\varepsilon^k h^{(k)}_{\boldsymbol{\nu},\gamma} = \sum_{\theta\in\Theta^o_{k,\boldsymbol{\nu},\gamma}} \mathrm{Val}(\theta) \tag{2.7}$$

for all $\boldsymbol{\nu}\in\mathbb{Z}^r$*, all* $k\in\mathbb{N}$ *and* $\gamma=1,\ldots,d$.

The expression (2.7) is well defined at fixed k and the sum over k gives what we call the *formal power series solution* for the equations for the parametric representation (2.1), (1.6) of the invariant tori.

3. The simplest resummation

The power series in ε in (2.1) and its Fourier transform defined by the sum over k of (2.7) may be not convergent *as a power series* (as far as we know). The problem is difficult because if in (2.7) we replace $\mathrm{Val}(\theta)$ with $|\mathrm{Val}(\theta)|$ the series certainly diverges.

Our aim, as stated in the introduction, is to show that nevertheless a meaning to the series can be given. We shall show that the tree values can be further decomposed into sums of several

other quantities and that the various contributions to the series can be rearranged by suitably collecting them into families: the sums of the contributions from each family leave us with a new series (no longer a power series in ε) which is in fact convergent and its sum solves the problem of constructing the parametric representations $\mathbf{h} = (\mathbf{a}, \mathbf{b})$, (2.1), of the invariant tori at least for all $\varepsilon \in \mathcal{E}$, with $\mathcal{E}$ a set with 0 as a density point (i.e. a Lebesgue point).

For this purpose we need to define and consider more involved trees and more involved definitions of their values. We begin by remarking that trees may contain *trivial nodes*, i.e. nodes with $\mathbf{0}$ harmonic separating two lines with equal momentum $\boldsymbol{\nu} \neq \mathbf{0}$.

One can suppose that *no tree contains trivial nodes* provided we use for all lines, with momentum $\boldsymbol{\nu} \neq \mathbf{0}$ and labels γ, γ' associated with the extremes, the *new* propagators

$$\overline{g}(x;\varepsilon) \stackrel{def}{=} (x^2 - M_0)^{-1}, \qquad x \stackrel{def}{=} \boldsymbol{\omega} \cdot \boldsymbol{\nu}, \qquad M_0 \stackrel{def}{=} \varepsilon \begin{pmatrix} 0 & 0 \\ 0 & \partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0) \end{pmatrix}. \tag{3.1}$$

This is a *resummation of many divergent series* obtained by adding the values of trees obtained from a tree without trivial nodes by "insertion" of an arbitrary number of trivial nodes on the branches with momentum $\boldsymbol{\nu} \neq \mathbf{0}$: this requires summing series, one per branch of a tree *without trivial nodes*, which are geometric series with ratio given by the $d \times d$ matrix $z = \frac{M_0}{(\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2}$; $|z|$ can be larger than 1 because the s non-zero eigenvalues εa_j, $j = 1, \ldots, s$, of M_0 are unrelated to $x = \boldsymbol{\omega}\cdot\boldsymbol{\nu}$.[1]

Therefore replacing $\sum_{p=0}^{\infty} z^p$ by $(1-z)^{-1}$ *is not rigorous and needs to be eventually justified.* Certainly we must at least suppose that $x^2 - M_0$ can be inverted: otherwise the values of the trees representing the new series might even be meaningless! (i.e. if some lines will have momentum $\boldsymbol{\nu}$ such that $\det(x^2 - M_0) = 0$). This happens for a dense set of ε's and we have to exclude such ε's by imposing conditions on the eigenvalues $\lambda^{[0]}_{r+j} \equiv \varepsilon a_j$, $j = 1, \ldots, s$, i.e. on ε.

For uniformity of notations it is convenient to assume that ε is in an interval $(\varepsilon_{\min}, 4\varepsilon_{\min}]$ related to the largest eigenvalue $\lambda^{[0]}_d \equiv a_s \varepsilon$ of M_0 by

$$\lambda^{[0]}_d \equiv \varepsilon\, a_s \in I_C \stackrel{def}{=} \Big(\frac{1}{4} C^2, C^2\Big], \qquad C \stackrel{def}{=} C_0 2^{-n_0}, \qquad n_0 \geq 0, \tag{3.2}$$

where C_0 is the Diophantine constant in (1.3) (fixed throughout the analysis); thus I_C is an interval of size $O(C^2)$ (i.e. $\frac{3}{4}C^2$). In other words we find it convenient to measure ε in units of $C_0^2 a_s^{-1}$ via an integer n_0. *We a priori assume, for simplicity, the restrictions* $a_s \varepsilon \leq C_0^2$ *and* $\varepsilon \leq 1$.

To give a meaning to $(x^2 - M_0)^{-1}$ it would suffice to require $|x^2 - \varepsilon a_j| \neq 0$ for all j thereby excluding "only" a denumerable (*dense*) set of values of ε, of 0 length; however stronger conditions will be needed in order to analyze the convergence problems and we begin by imposing them in a form which will be useful later. Setting for later use $\underline{\lambda}^{[0]}_j(\varepsilon) \stackrel{def}{=} \lambda^{[0]}_j$, the conditions that we impose on $\lambda^{[0]}_j$, i.e. on ε, are that for all $x = \boldsymbol{\omega} \cdot \boldsymbol{\nu} \neq 0$ and for all independent choices of the signs $+$ or $-$

$$\Gamma(x) \stackrel{def}{=} \min_{j \geq i} \Big\{ \Big| |x| - \sqrt{\underline{\lambda}^{[0]}_j(\varepsilon)} \Big|, \ \Big| x \pm \sqrt{\underline{\lambda}^{[0]}_j(\varepsilon)} \pm \sqrt{\underline{\lambda}^{[0]}_i(\varepsilon)} \Big| \Big\} \geq 2^{-(\overline{n}_0 - 1)/2} \frac{C_0}{|\boldsymbol{\nu}|^{\tau_1}} \tag{3.3}$$

[1] Note that since the tree lines are numbered (i.e. they are regarded as distinct) adding p nodes on a line ℓ changes the combinatorial factor $k!^{-1}$ in (2.5) into $(k+p)!^{-1}$: however the new p lines thus produced can be chosen in $\binom{k+p}{p}$ ways and ordered in $p!$ ways so that we can ignore the extra number labels on ℓ and use as combinatorial factor $(k+p)!^{-1}\binom{k+p}{p} p! = k!^{-1}$.

for τ_1 suitably large and $\overline{n}_0$ suitably larger than n_0, see (4.2). This excludes a closed set of values of ε in the considered interval I_C, (3.2): its measure can be estimated without difficulties. Let

$$\tau_1 = \tau_0 + r + 1, \tag{3.4}$$

be a convenient, although somewhat arbitrary, choice; then the total measure of the excluded set is

$$\le \, 2^{-(\overline{n}_0 - 1)/2} C^2 \, K, \tag{3.5}$$

where K is a suitable constant; see Appendix A2. Hence the measure of the complement of the set $\mathcal{E}_{\overline{n}_0 - 1}$ where (3.3) is verified is a *small fraction* of order $C^{1/2}$ of the measure of the interval I_C, whose size is proportional to C^2, in which we let ε vary, at least if $\overline{n}_0$ is large.

4. Resummations: semantic and heuristic considerations

Replacing the propagators x^{-2} of the lines by $(x^2 - M_0)^{-1}$ we obtain a representation of the parametric equations $\mathbf{h}$ involving simpler trees (i.e. trees with no trivial nodes). The new representation is a series in which each term is well defined if ε is in the large set $\mathcal{E}_{\overline{n}_0 - 1} \subset I_C$ in which (3.3) holds. This is quite different from the original Lindstedt series in (2.7) whose terms are well defined for all ε.

We should also stress that the resummed series is in a sense more natural: the $\mathbf{0}$ momentum lines now appear as less anomalous because their propagator is much more closely related to $(x^2 - M_0)^{-1}$. One can say that it is just the latter evaluated at $x = 0$ with the meaningless entries (i.e. the first r diagonal entries) replaced by 0. Another way of saying the latter property is that lines ℓ with $\mathbf{0}$ momentum and labels $\gamma_\ell, \gamma'_\ell \le r$ are forbidden. One should not be surprised by this fact: it is the generalization of the corresponding property in the case of maximal tori ($r = d$) in which this means that lines with $\mathbf{0}$ momentum are forbidden. The latter property goes back to Poincaré's theory of the Lindstedt series and is the key to the proof of the KAM theorem and of cancellations which make the formal Lindstedt series for maximal tori absolutely convergent; see Refs. [E1] and [Ga]. However the new series is still only a formal representation because it is by no means clear that it is absolutely converges.

The next natural idea is to try to establish convergence by further modifying the propagators, changing at the same time the trees structure, until one achieves a formal representation whose convergence will be "easy" to check. Once achieved a formal representation which is convergent we shall have to check that it really solves the equations for $\mathbf{h}$.

The modification of the trees structure will be performed by steps: at each step, labeled by an integer $n = 0, 1, \ldots$, the propagators of the lines with non-zero momentum will have been modified acquiring labels $[0], [1], \ldots [n-1]$, or the label $[\ge n]$, indicating that they are given no longer by $(x^2 - M_0)^{-1}$ but by a matrix proportional to $(x^2 - \mathcal{M}^{[\le p]})^{-1}$, if their label is $[p]$, with $p < n$, or (with a different proportionality factor) to $(x^2 - \mathcal{M}^{[\le n]})^{-1}$, if their label is $[\ge n]$; here $\mathcal{M}^{[\le p]}$ are suitable matrices. *Here and in the following the symbols $[\le n]$ and $[\ge n]$ are consistently used. Hence $[\ge n]$ does not denote the set of scales $[p]$ with $p \ge n$, and in fact it is just a different scale; likewise $[\le n]$ does not "include" $[p]$ even if $p \le n$. In other words one has to regard the symbols $[\le n]$, $[n]$ and $[\ge n]$ as unrelated symbols. This might appear unusual but it turns out to be a good notation for our purposes.*

The proportionality factor depends on x and contains cut-off functions which vanish unless $x^2 - \mathcal{M}^{[\le p]}$ has smallest eigenvalue of order $O(2^{-2p} C_0^2)$; the cut-offs are so devised that if the

propagator does not vanish its denominator has a minimum size proportional to 2^{-2p} and the ratio between its minimum and maximum values will be bounded above and below by a *p-independent constant.* No modification will be made of the propagators of the $\mathbf{0}$ momentum lines: for uniformity of notation we shall attach a label $[\infty]$ to such lines.

Considering trees with no trivial nodes in which each line carries also an extra *scale* label $[\infty],[0],[1],\ldots[n-1],[\ge n]$ a new formal representation of $\mathbf{h}$ will be obtained by assigning, to the trees, values defined by the same formula in (2.5), with the propagators G_ℓ replaced by the new propagators, that we denote $g_\ell^{[p]}$ if the line ℓ carries the label $[p]$, with $p=\infty,0,\ldots,n-1$, and $g_\ell^{[\ge n]}$ if the line carries the label $[\ge n]$. When the line ℓ is on scale $[p]$, with $p=0,\ldots,n-1$, or $[\ge n]$ or $[\infty]$, then the corresponding propagator will be proportional to $(x^2-\mathcal{M}^{[\le p]})^{-1}$ or $(x^2-\mathcal{M}^{[\le n]})^{-1}$ or $(\varepsilon\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0))^{-1}$.

The construction will be performed in such a way that the matrices $(x^2-\mathcal{M}^{[\le p]})$ will be defined by series which *will be proved to be convergent*; furthermore if we only considered the contributions to the formal representation of $\mathbf{h}$ coming from trees *in which no propagator carries the "last label"* $[\ge n]$ *then the corresponding series would be convergent.*

We express the latter property by saying that *the performed resummations regularize the formal representation of* $\mathbf{h}$ *down to scale* $[n-1]$*, or that the propagators singularities are probed down to scale* $[n-1]$. The problem of course remains to understand the contributions from the trees containing lines with label $[\ge n]$. The construction will be such that their propagators are also properly defined because the matrices $\mathcal{M}^{[\le n]}$ will always be well defined by convergent series (as we shall see). However for the lines whose label is $[\ge n]$ no useful positive lower bound, not even n-dependent, can be given on the smallest eigenvalue of the denominators in the corresponding propagators.

We shall say that the lines with scale $[\ge n]$ probe the singularity all the way down to the smallest frequencies or all the way down in the *infrared* scales. Thus in spite of the convergence of the contributions to $\mathbf{h}$ coming from trees with labels $[\infty],[0],[1],\ldots,[n-1]$ the representation of $\mathbf{h}$ remains formal.

Therefore we shall proceed by increasing the value of n trying to take the limit $n\to\infty$. This is the procedure followed in the case of the theory of hyperbolic tori in Ref. [GG]. In that case, however, the propagators denominators $(x^2-\mathcal{M}^{[\le n]})$ had eigenvalues *always bounded below proportionally to* x^2. Indeed the last s eigenvalues of $\mathcal{M}^{[\le n]}$ were negative whereas the first r remained close to zero within $O(\varepsilon x^2)$ (a non-trivial property, however, due to remarkable cancellations well known in the KAM theory, [Ga]).

Here the matrices $x^2-\mathcal{M}^{[\le n]}$ will be shown to have the first r eigenvalues differing by a factor $(1+O(\varepsilon^2))$ and the last s differing by $O(\varepsilon^2)$ with respect to those of x^2-M_0 (which has by construction r eigenvalues x^2 and s eigenvalues $x^2-\varepsilon a_j$ $j=1,\ldots,s$). Thus the denominators can become small because *either* x^2 gets close to 0 *or* because it gets close to $\varepsilon a_1,\ldots\varepsilon a_s$. Therefore the regularization will have to be split in two parts. The first part will concern regularizing the scales $[p]$ with p such that the eigenvalues of $x^2-\mathcal{M}^{[\le p]}$ remains bounded below proportionally to x^2; we shall call this part of the analysis the *high frequencies resummation.* The other part, which we shall call the *infrared resummation*, will concern the regularization of the scales $[p]$, in which x can be so close to some εa_j that the denominators cannot be bounded below proportionally to x^2.

We associate with each momentum $\boldsymbol{\nu}$ the frequency $x=\boldsymbol{\omega}\cdot\boldsymbol{\nu}$ and we measure the strength of

this resonance by the integer p if $D(x;\varepsilon)\simeq C_0^2 2^{-2p}$, with

$$D(x;\varepsilon)=\min_j\Big|x^2-\underline{\lambda}_j^{[0]}(\varepsilon)\Big|\stackrel{def}{=}\Big|x^2-\underline{\lambda}^{[0]}_{j_\varepsilon(x)}(\varepsilon)\Big|. \tag{4.1}$$

Therefore the condition that the resonance strength of the frequency x be bounded below proportionally to $|x|$ is that p is not too large compared to n_0 defined in (3.2), so that x^2 stays away from the corresponding eigenvalue $\underline{\lambda}_j^{[0]}(\varepsilon)$ by more than a small fraction of the minimum separation δ between the distinct eigenvalues. For instance we can require $D(x;\varepsilon)\ge 2^{-2(\overline{n}_0+1)}C_0^2\ge\delta/4$. This gives $p\le\overline{n}_0$, with

$$\overline{n}_0=n_0+\overline{n},\qquad \overline{n}\stackrel{def}{=}-1+\frac12\log_2\frac1\rho,\qquad \rho=\frac14 a_s^{-1}\min\{a_1,\min_j\{a_{j+1}-a_j\}\}. \tag{4.2}$$

In fact the requirement could be fulfilled with $\overline{n}$ one unit larger: the interest of using the above value of $\overline{n}$ will emerge later (if $s=1$ one interprets $\rho=\frac14$).

We then perform the analysis by defining recursively the matrices $\mathcal{M}^{[\le p]}(x;\varepsilon)$ for $p=0,\ldots,\overline{n}_0$ with eigenvalues $\lambda_j^{[p]}(x,\varepsilon)$ verifying for a suitable constant $\gamma>0$

$$|\lambda_j^{[p]}(x,\varepsilon)-\lambda_j^{[0]}(\varepsilon)|<\gamma\varepsilon^2,\qquad p\le\overline{n}_0, \tag{4.3}$$

so that if the label p of the line with frequency x is $p\le\overline{n}_0$ then one has, if $\frac12 a_s 2^{-2(\overline{n}+1)}-\gamma\varepsilon\ge0$,

$$|x^2-\lambda_j^{[p]}(x,\varepsilon)|\ge\frac12 D(x,\varepsilon)+\frac12 D(x,\varepsilon)-\gamma\varepsilon^2\ge\frac12 D(x,\varepsilon)\ge 2^{-2(\overline{n}+2)}|x|^2 \tag{4.4}$$

where the last step is obvious if $|x|^2\ge 2\underline{\lambda}_d^{[0]}(\varepsilon)$, otherwise it follows from the inequality

$$D(x;\varepsilon)\ge 2^{-2(\overline{n}+1)}2^{-2n_0}C_0^2\ge 2^{-2(\overline{n}+1)}\underline{\lambda}_d^{[0]}(\varepsilon)\ge 2^{-2(\overline{n}+1)-1}|x|^2. \tag{4.5}$$

We can say that for $p\le\overline{n}_0$ the strength of the singularity is dominated by the distance $|x|$ to the origin, i.e. by the "classical" small divisors x^{-2} provided, of course, the matrices $x^2-\mathcal{M}^{[\le p]}(x;\varepsilon)$ remain close enough to x^2-M_0 (which we shall check). Furthermore the convergence of the sum of all values of trees with no line label $[\ge\overline{n}_0]$ will be performed exactly along the lines of Ref. [GG] because the bound (4.4) guarantees that in evaluating such trees one does not probe the singularities close to the eigenvalues of M_0.

The departure from the method in Ref. [GG] occurs when we consider trees in which lines bear the label $[\ge\overline{n}_0]$. The problem will again be studied by a multiscale analysis which will have to be suitably modified to allow probing the new singularities arising from the resonances between the frequencies x and the $\sqrt{\underline{\lambda}_j^{[0]}(\varepsilon)}$, $j>r$. The difficulty is that the propagator $g_\ell^{[\ge\overline{n}_0]}$ *will not be singular exactly at the frequencies* $\sqrt{\underline{\lambda}_j^{[0]}(\varepsilon)}\ne0$ but at the frequencies fixed by the roots of the eigenvalues $\lambda_j^{[\le\overline{n}_0]}(x;\varepsilon)$ of the matrices $\mathcal{M}^{[\le\overline{n}_0]}(x;\varepsilon)$. The latter not only are slightly different from those of M_0 but will turn out to depend also on x.

This means that $D(x;\varepsilon)$ and even $\Delta(x;\varepsilon)=\min_j\big|x^2-\lambda_j^{[\le\overline{n}_0]}(x;\varepsilon)\big|$ no longer provide a good estimate of the strength of the singularity, because D,Δ vanish at the "wrong places". In fact we

shall have to perform a multiscale analysis to resolve the infrared singularities, and it will happen that at each of the new scales with labels $[p]$, with $p \geq \overline{n}_0$, the singularities will keep moving.

Suppose to have regularized the series up to scale $[n-1]$, with $n > \overline{n}_0$, introducing suitably matrices $\mathcal{M}^{[\leq p]}(x;\varepsilon)$, with $p = \overline{n}_0, \ldots, n$, thus pushing the probe of the singularities down to scales $C_0 2^{-n}$; then to avoid meaningless expressions we shall have to impose on the eigenvalues of the last propagator, proportional to $(x^2 - \mathcal{M}^{[\leq n]}(x;\varepsilon))^{-1}$, a condition like (3.3). Since the eigenvalues depend on n and x this risks to imply that we have to discard too many ε's; in the limit $n \to \infty$: when, finally, the singularities will have been probed on all scales, or even for large enough scales, we might be left with an empty set of ε's rather than with a set of almost full measure.

Physically the difficulty shows up because of the possibility of resonances between the proper frequencies of the quasi-periodic motion on the tori and the normal frequencies. It will be studied and solved in Section 6 below, while in Section 5 we shall discuss the simpler regularization of the series for $\mathbf{h}$ on the high frequency scales.

The spirit informing the analysis is very close to the techniques used in harmonic analysis, in quantum field theory and in statistical mechanics, known as "renormalization group methods" (see Refs. [Fe], [GM1], [Ga01] and [Ga02]). The methods are also based on a "multiscale decomposition" of the propagators singularities. We introduced and adopt the above terminology because we feel that it is suggestive and provides useful intuition at least to the readers who have some acquaintance with the renormalization group approach and multiscale analysis.

5. Non-resonant resummations

The resummations will be defined via trees with no trivial nodes and with lines bearing further labels. Moreover the definition of propagator will be changed, hence the values of the trees will be different from the ones in Section 3: they are constructed recursively.

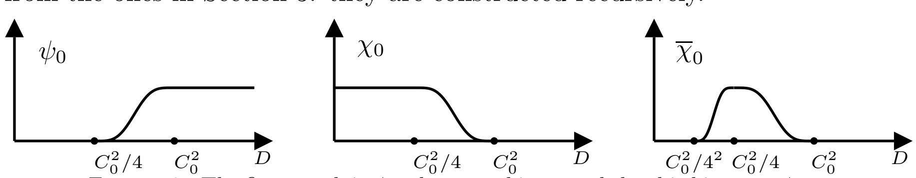

FIGURE 2. The first graph is ψ_0, the second is χ_0 and the third is $\overline{\chi}_0 \equiv \psi_1\chi_0$.

Instead of the sharp multiscale decomposition considered in Ref. [GG] here it will be convenient to work with a smooth one as in Ref. [Ge]. Let $\psi(D)$ be a C^∞ non-decreasing compact support function defined for $D \geq 0$, see Fig. 2, such that

$$\psi(D) = 1, \quad \text{for} \quad D \geq C_0^2, \qquad \psi(D) = 0, \quad \text{for} \quad D \leq C_0^2/4, \tag{5.1}$$

where C_0 is the Diophantine constant in (1.3), and let $\chi(D) = 1 - \psi(D)$. Define also $\psi_n(D) = \psi(2^{2n}D)$ and $\chi_n(D) = \chi(2^{2n}D)$ for all $n \geq 0$. Hence $\psi_0 = \psi$, $\chi_0 = \chi$ and

$$1 \equiv \psi_n(\Delta(x;\varepsilon)) + \chi_n(\Delta(x;\varepsilon)), \qquad \text{for all } n \geq 0, \tag{5.2}$$

for all choices of the function $\Delta(x;\varepsilon) \geq 0$: in particular for $\Delta(x,\varepsilon) = D(x)$ with $D(x)$ defined in (5.3) below. We set the following notations.

DEFINITION 1. *Let $\overline{n}_0, \overline{n}$ be as in (4.2) and $D(x;\varepsilon)$ as in (4.1).*

(i) Divide the interval $I_C \equiv [\varepsilon_{\min}, 4\varepsilon_{\min}]$*, where* ε *varies, see (3.2), into a finite number of small intervals* I *of size smaller than* $\frac{1}{2}\varepsilon_{\min}\rho$*, see (4.2), i.e. smaller than a fraction of the minimum separation between the eigenvalues* $0, a_1, \ldots, a_s$*. Define*

$$D(x;I) = \min_{\varepsilon\in I} D(x;\varepsilon) = \min_{\varepsilon\in I}\min_{j}\left|x^2 - \underline{\lambda}_j^{[0]}(\varepsilon)\right| = \min_{\varepsilon\in I}\left|x^2 - \underline{\lambda}_{j(x)}^{[0]}(\varepsilon)\right|. \tag{5.3}$$

where $j(x)$ *is the smallest value of* j *for which the last equality holds: exceptionally there might be 2 such labels. The* $j(x)$ *is* ε*-independent, by construction, for* $\varepsilon \in I$*.*

Remarks. (1) Note that, as a consequence of the definition of the intervals I and of $D(x;I)$ as given by (5.3), one has, for all $\varepsilon \in I$,

$$\min_j \left|x^2 - \underline{\lambda}_j^{[0]}(\varepsilon)\right| \ge \frac{1}{2}\left|x^2 - \underline{\lambda}_{j(x)}^{[0]}(\varepsilon)\right|, \tag{5.4}$$

(2) If ε is in one of the intervals I and x verifies $D(x;I) \le C_0^2 2^{-2\overline{n}_0}$ then there is only one value of j for which last equality in (5.3) holds.
(3) *We shall fix, from now on,* ε *in one of the intervals* $I \subseteq I_C$. Remark that $D(x;I)$ is piecewise linear in x^2 with slope equal to 1 in absolute value for x in the regions where it will be considered (see below) and we simplify the notation by setting

$$D(x) \stackrel{def}{=} D(x;I). \tag{5.5}$$

(4) The number of intervals $I \subset I_C$ can and will be taken independent of $\varepsilon_{\min}$, i.e. of the interval I_C where ε varies, and equal to a fixed integer $\le 6\rho^{-1}$.
(5) From now on we only consider trees with no trivial nodes.

A simple way to represent the value of a tree as sum of many terms is to make use of the identity in (5.2). The propagator $g(x;\varepsilon) \equiv g^{[\ge 0]}(x;\varepsilon) \stackrel{def}{=} (x^2 - M_0)^{-1}$ of each line with non-zero momentum (hence with $x \ne 0$) is written as

$$g^{[\ge 0]}(x;\varepsilon) = \psi_0(D(x))\, g^{[\ge 0]}(x;\varepsilon) + \chi_0(D(x))\, g^{[\ge 0]}(x;\varepsilon) \stackrel{def}{=} g^{[0]}(x;\varepsilon) + g^{\left\{\ge 1\right\}}(x;\varepsilon), \tag{5.6}$$

and we note that $g^{[0]}(x;\varepsilon)$ vanishes if $D(x)$ is smaller than $(C_0/2)^2$, see Fig. 2.

If we replace each $g^{[\ge 0]}(x;\varepsilon)$ with the sum in (5.6) then the value of each tree of order k is split as a sum of up to 2^k terms[2] which can be identified by affixing on each line with momentum $\boldsymbol{\nu} \ne \mathbf{0}$ a label [0] or $\left\{\ge 1\right\}$. Further splittings of the tree values can be achieved as follows.

Definition 2. *For* $p = 1, \ldots, \overline{n}_0$*, let* $\mathcal{M}^{[p]}(x;\varepsilon)$ *be matrices with eigenvalues* $\lambda_j^{[p]}(x;\varepsilon)$*,* $p = 1, \ldots, n$*; we set* $\mathcal{M}^{[0]}(x;\varepsilon) \equiv M_0$ *and* $\mathcal{M}^{[\le n]}(x;\varepsilon) = \sum_{p=0}^{n} \mathcal{M}^{[p]}(x;\varepsilon)$*. Define for* $0 < n \le \overline{n}_0 - 1$

$$\begin{aligned}
g^{[n]}(x;\varepsilon) &\stackrel{def}{=} \frac{\psi_n(D(x))\prod_{m=0}^{n-1}\chi_m(D(x))}{x^2 - \mathcal{M}^{[\le n]}(x;\varepsilon)},\\
g^{\left\{\ge n\right\}}(x;\varepsilon) &\stackrel{def}{=} \frac{\prod_{m=0}^{n-1}\chi_m(D(x))}{x^2 - \mathcal{M}^{[\le n-1]}(x;\varepsilon)},\\
g^{[\ge n]}(x;\varepsilon) &\stackrel{def}{=} \frac{\prod_{m=0}^{n-1}\chi_m(D(x))}{x^2 - \mathcal{M}^{[\le n]}(x;\varepsilon)},
\end{aligned} \tag{5.7}$$

[2] Not necessarily 2^k because there might be lines on scale $[\infty]$ whose propagator is not decomposed.

and $g^{[0]}(x;\varepsilon) = \psi_0(D(x))\,(x^2 - M_0)^{-1}$. *We call the labels* $[n], \{\geq n\}, [\geq n]$ *scale labels.*

Remarks. (1) The products $\prod_{m=0}^{n-1} \chi_m(D(x))$ can be simplified to involve only the last factor: we keep the notation above as it is a notation that naturally reflects the construction. The propagators $g^{\{\geq n\}}$ play a subsidiary role and are here for later reference.
(2) The matrices $\mathcal{M}^{[p]}(x;\varepsilon)$ will be defined recursively under the requirement that the functions $\mathbf{h}$ defining the parametric equations of the invariant torus will be expressed in terms of trees whose lines carry scale labels indicating that their values are computed with the propagators in (5.7).
(3) Note that if we defined $\mathcal{M}^{[\leq p]}(x;\varepsilon) \equiv M_0$, i.e. $\mathcal{M}^{[p]} \equiv 0$ for $p > 0$, then (recall that we consider only trees without trivial nodes) we would naturally decompose (see below for details) the tree values into sums of many terms keeping obviously each total sum constant by repeatedly using (5.2), thus meeting the requirement in Remark (2) above. *This would be of no interest.* Therefore we shall try to define the matrices $\mathcal{M}^{[p]}(x;\varepsilon)$ so that the sum of the values of *new trees* (with no trivial nodes and whose nodes and lines ℓ still carry harmonic and momentum labels as well as scale labels $[\infty], [0], \ldots, [n-1], [\geq n]$) remain the same provided their values are evaluated by using the propagators in (5.7) and *we shall try to define* $\mathcal{M}^{[\leq p]}(x;\varepsilon)$*, so that there is also control of the convergence.*
(4) *In other words we try to obtain a graphical representation of* $\mathbf{h}$*, involving values of trees which are easier to study at the price of needing more involved propagators.* This is a typical method employed in KAM theory [GBG], and in other fields.

To define recursively the matrices we introduce the notions of clusters and of self-energy clusters of a tree whose lines and nodes carry the same labels introduced so far and *in addition* each line carries a scale label which can be either $[\infty]$, if the momentum of the line is zero, or $[p]$, with $p = 0, \ldots, \overline{n}_0 - 1$, or $[\geq \overline{n}_0]$. Given a tree θ decorated in this way we give the following definition, for $n \leq \overline{n}_0$.

Definition 3. (Clusters)
(i) A cluster T *on scale* $[n]$*, with* $0 \leq n$*, is a maximal set of nodes and lines connecting them with propagators of scales* $[p]$*,* $p \leq n$*, one of which, at least, of scale exactly* $[n]$*. We denote with* $V(T)$ *and* $\Lambda(T)$ *the set of nodes and the set of lines, respectively, contained in* T*. The number of nodes in* T *will define the order of* T*, and it will be denoted with* k_T*.*
(ii) The $m_T \geq 0$ *lines entering the cluster* T *and the possible line coming out of it (unique if existing at all) are called the* external lines *of the cluster* T*.*
(iii) Given a cluster T *on scale* $[n]$*, we shall call* $n_T = n$ *its scale.*

Remarks. (1) For instance if $n = 0$ the scale of the lines in the cluster can only be $[0]$.
(2) Here $n \leq \overline{n}_0 - 1$. However the definition above is given in such a way that it will extend unchanged when also scales larger than $\overline{n}_0$ will be introduced.
(3) The clusters of a tree can be regarded as sets of lines hierarchically ordered by inclusion and have hierarchically ordered scales.
(4) A cluster T is not a tree (in our sense); however we can uniquely associate a tree with it by adding the entering and the exiting lines and by imagining that the lower extreme of the exiting line is the root and that the highest extremes of the entering lines are nodes carrying a harmonic

label equal to the momentum flowing into them.

Definition 4. (Self-energy clusters)
(i) We call self-energy cluster *of a tree θ any cluster T of scale $[n]$ such that T has only one entering line ℓ_T^2 and one exiting line ℓ_T^1, and furthermore $\sum_{\mathfrak{v}\in V(T)} \boldsymbol{\nu}_{\mathfrak{v}} = \mathbf{0}$.*
(ii) The degree of a self-energy cluster is the number of nodes.

Remark. The essential property of a self-energy cluster is that it has necessarily just one entering line and one exiting line, and both have *equal momentum* (because $\sum_{\mathfrak{v}\in V(T)} \boldsymbol{\nu}_{\mathfrak{v}} = \mathbf{0}$). Note that both scales of the external lines of a self-energy cluster T are strictly larger than the scale of T as a cluster, but they can be different from each other by at most one unit. Furthermore the degree of a self-energy cluster is ≥ 2. Of course no self-energy cluster can contain any line on scale $[\infty]$.

Definition 5. (Self-energy matrices)
(i) Let $\Theta^{\mathcal{R}}_{k,\boldsymbol{\nu},\gamma}$ be the set of trees of order k with root line momentum $\boldsymbol{\nu}$ and root label γ which contain neither self-energy clusters nor trivial nodes. Such trees will be called renormalized trees.
(ii) We denote with $\mathcal{S}^{\mathcal{R}}_{k,n}$ the set of self-energy clusters of order k and scale $[n]$ which do not contain any other self-energy cluster nor any trivial node; we call them renormalized self-energy clusters *on scale n.*
(iii) Given a self-energy cluster $T \in \mathcal{S}^{\mathcal{R}}_{k,n}$ we shall define the self-energy value of T as the matrix[3]

$$\mathcal{V}_T(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon) = \frac{\varepsilon^k}{(k-1)!}\Big(\prod_{\ell\in\Lambda(T)} g_\ell^{[n_\ell]}\Big)\Big(\prod_{\mathfrak{v}\in V(T)} F_{\mathfrak{v}}\Big), \tag{5.8}$$

where $g_\ell^{[n_\ell]} = g^{[n_\ell]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu}_\ell;\varepsilon)$. Note that, necessarily, $n_\ell \leq n$. The $k_T - 1$ lines of the self-energy cluster T will be imagined as distinct and to carry a number label ranging in $\{1,\ldots,k_T-1\}$.

The recursive definition of the matrices $\mathcal{M}^{[n]}(x;\varepsilon)$, $n \geq 1$, will be (if the series converges)

$$\mathcal{M}^{[n]}(x;\varepsilon) = \Big(\prod_{p=0}^{n-1}\chi_p(D(x))\Big)\sum_{k=2}^{\infty}\sum_{T\in\mathcal{S}^{\mathcal{R}}_{k,n-1}}\mathcal{V}_T(x;\varepsilon) \stackrel{def}{=} \Big(\prod_{p=0}^{n-1}\chi_p(D(x))\Big)M^{[n]}(x;\varepsilon), \tag{5.9}$$

where the self-energy values are evaluated by means of the propagators on scales $[p]$, with $p = 0,\ldots,n$, which makes sense because we have already defined the propagators on scale $[0]$ and the matrices $\mathcal{M}^{[0]}(x;\varepsilon) \equiv M_0$ (cf. Definition 2).

With the above new definitions we have the formal identities

$$h_{\boldsymbol{\nu},\gamma} = \sum_{k=1}^{\infty}\sum_{\theta\in\Theta^{\mathcal{R}}_{k,\nu,\gamma}} \mathrm{Val}(\theta), \tag{5.10}$$

where we have redefined the *value* of a tree $\theta \in \Theta^{\mathcal{R}}_{k,\boldsymbol{\nu},\gamma}$ as

$$\mathrm{Val}(\theta) = \frac{\varepsilon^k}{k!}\Big(\prod_{\ell\in\Lambda(\theta)} g^{[\eta_\ell]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu}_\ell;\varepsilon)\Big)\Big(\prod_{\mathfrak{v}\in V(\theta)} F_{\mathfrak{v}}\Big), \tag{5.11}$$

[3] This is a matrix because the self-energy cluster inherits the labels γ,γ' attached to the endnode of the entering line and to the initial node of the exiting line.

with $[\eta_\ell] = [\infty], [0], \ldots, [\overline{n}_0 - 1], [\geq \overline{n}_0]$. Note that (5.10) is not a power series in ε.

The statement in (5.10) requires some thought, but it turns out to be a tautology, see also Ref. [GG], and Ch. VIII in Ref. [GBG], *if one neglects convergence problems* which, however, will occupy us in the rest of this paper. A sketch of the argument is as follows.

Imagine that we have only scales $[\infty], [0], \ldots, [n-1], [\geq n]$, i.e. we have performed the scale decomposition of the propagators up to scale $[n-1]$ and we have not decomposed the propagators on scale $[\geq n]$ and that we have checked the statement (5.9) and (5.10) (trivially true for $n = 0$).

Given a tree $\theta \in \Theta^{\mathcal{R}}_{k,\boldsymbol{\nu},\gamma}$ with lines carrying labels $[p]$ with $p = 0, \ldots, n-1$ or $[\geq n]$ or $[\infty]$, we can split the propagators $g^{[\geq n]}(x;\varepsilon)$ as $g^{[n]}(x;\varepsilon) + g^{\left\{\geq n+1\right\}}(x;\varepsilon)$ as in (5.6) with $g^{[n]}(x;\varepsilon) = \psi_n(D(x)) g^{[\geq n]}(x;\varepsilon)$ and $g^{\left\{\geq n+1\right\}}(x;\varepsilon) = \chi_n(D(x)) g^{[\geq n]}(x;\varepsilon)$. In this way we get new trees *which in general contain self-energy clusters of scale* $[n]$. We can in fact construct infinitely many trees with self-energy clusters of scale $[n]$ simply by *inserting* an arbitrary number of them on any line ℓ with scale $\{\geq n+1\}$.

The values of the trees obtained by $k \geq 0$ such self-energy *insertions* on a given line of a tree in $\Theta^{\mathcal{R}}_{k,\boldsymbol{\nu},\gamma}$ can be arranged into a geometric progression: in fact they differ only by a factor equal to the value of the integer power k in $g^{\left\{\geq n+1\right\}}(x;\varepsilon)\big(M^{[n+1]}(x;\varepsilon) g^{\left\{\geq n+1\right\}}(x;\varepsilon)\big)^{k+1}$ if $M^{[n+1]}(x;\varepsilon)$ is defined as in (5.9), where the $\mathcal{V}_T(x;\varepsilon)$ are evaluated by using as propagators $g^{[p]}(x;\varepsilon)$, with $0 \leq p \leq n$ or $p = \infty$, for the lines carrying a scale label $[p]$. *Summation over k will simply change $g^{\left\{\geq n+1\right\}}(x;\varepsilon)$ into $g^{[\geq n+1]}(x;\varepsilon)$ and at the same time one shall have to consider only trees with no self-energy cluster of scale $[n]$ nor of scale $[p]$ with $p < n$ and with lines carrying scale labels $[\infty], \ldots, [n]$ or $[\geq n+1]$.* In this way we prove (5.10) for all $n \leq \overline{n}_0$: we could continue, but for the reasons outlined in Section 4, we decide to stop the resummations at this scale.

In other words the above is a generalization of the simple resummation considered in Section 3. The result is still *as formal as the Lindstedt series we started with* even assuming convergence of the series in (5.9). In fact the consequent expression for $\mathbf{h}$ cannot even be, if taken literally, correct because as in Section 3 *the denominators in the new expressions could even vanish because no lower cut-off operates on the lines with scale $[\geq \overline{n}_0]$ as the third of (5.7) shows.*

To proceed we must first check that the series (5.9) defining $M^{[n]}(x;\varepsilon)$ are really convergent. In spite of the last comment this will be true because in the evaluation of $M^{[n]}(x;\varepsilon)$ *the only propagators needed have scales $[p]$ with $p \leq n-1$* so that, see the factors $\psi_n(D(x)), \chi_n(D(x))$ in (5.7), their denominators not only do not vanish but have controlled sizes that can be bounded below proportionally to x^2 by (4.4), i.e. simply by a constant times $C_0^2 |\boldsymbol{\nu}|^{-2\tau_0}$, see (1.3).

In Ref. [GG] it has been shown *by a purely algebraic symmetry argument* that, as long as one can prove convergence of the series in (5.9), the matrices $M^{[n]}(x;\varepsilon)$ are Hermitian and $(M^{[n]}(x;\varepsilon))^T = M^{[n]}(-x;\varepsilon)$. Furthermore we should expect that the eigenvalues of the matrix $\mathcal{M}^{[\leq n]}(x;\varepsilon)$ should be approximately located either near 0 or near $\varepsilon a_1, \ldots, \varepsilon a_s$ at least within $O(\varepsilon^2)$.

The expectation relies on Ref. [GG] (see Eq. (3.25)) where the following "*cancellations result*" is derived for n_0 large enough (hence for ε small because $2^{-2n_0-2} < \varepsilon a_s \leq 2^{-2n_0} C_0^2$): we reproduce the proof in Appendix A3 below, adapting it to the present notations.

Lemma 2. *If n_0 is large enough and $n \leq \overline{n}_0 = n_0 + \overline{n}$ (see (4.2)) then the following properties hold.*

(i) The matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$, $x=\boldsymbol{\omega}\cdot\boldsymbol{\nu}$, are Hermitian and can be written as

$$\mathcal{M}^{[\le n]}(x;\varepsilon)=\begin{pmatrix}\mathcal{M}^{[\le n]}_{\alpha\alpha}(x;\varepsilon) & \mathcal{M}^{[\le n]}_{\alpha\beta}(x;\varepsilon)\\ \mathcal{M}^{[\le n]}_{\beta\alpha}(x;\varepsilon) & \mathcal{M}^{[\le n]}_{\beta\beta}(x;\varepsilon)\end{pmatrix} \tag{5.12}$$

where the labels α run over $\{1,\ldots,r\}$ and β over $\{r+1,\ldots,s\}$.
(ii) One has $\mathcal{M}^{\le[n]}(x;\varepsilon)=(\mathcal{M}^{[\le n]}(-x;\varepsilon))^T$, so that the eigenvalues of $\mathcal{M}^{[\le n]}(x;\varepsilon)$ verify the symmetry property $\lambda^{[n]}_j(x;\varepsilon)=\lambda^{[n]}_j(-x,\varepsilon)$, i.e. they are functions of x^2. [4]
(iii) Let $\partial^{\pm}_x$ be right and left x-derivatives, then

$$\begin{aligned}
&\|\mathcal{M}^{[n]}(x,\varepsilon)\|\le B\,\varepsilon^2\,e^{-\kappa_1 2^{n/\tau}},\quad \|\partial^{\pm}_x\mathcal{M}^{[\le n]}(x,\varepsilon)\|\le B\varepsilon^2 a_s^{-1/2},\quad \|\partial^{\pm}_\varepsilon\mathcal{M}^{[\le n]}(x,\varepsilon)\|\le B\,\varepsilon,\\
&\|\mathcal{M}^{[n]}_{\alpha\alpha}(x;\varepsilon)\|\le B\,e^{-\kappa_1 2^{n/\tau}}\,\min\{\varepsilon^2,\varepsilon\,x^2 a_s^{-1}\},\\
&\|\mathcal{M}^{[n]}_{\alpha\beta}(x;\varepsilon)\|\le B\,e^{-\kappa_1 2^{n/\tau}}\,\min\{\varepsilon^2,\varepsilon^{\frac32}|x|\,a_s^{-1/2}\},\\
&\|\mathcal{M}^{[n]}_{\beta\beta}(x;\varepsilon)\|\le B\,e^{-\kappa_1 2^{n/\tau}}\,\varepsilon^2,
\end{aligned} \tag{5.13}$$

for $n\le\overline{n}_0$ and for suitable $\overline{n}_0$-independent constants $B,\kappa_1,\tau>0$; one can take $\tau=\tau_0$.

While κ_1 is dimensionless the constants A',A,B have same dimension (of a frequency square): this is the purpose of introducing appropriate powers of a_s.

General properties of matrices and (5.13) imply, see Appendix A4,

$$\begin{aligned}
&A'<|\partial_\varepsilon\lambda^{[n]}_j(x;\varepsilon)|<A, \qquad a_s^{\frac12}|\partial^{\pm}_x\lambda^{[n]}_j(x;\varepsilon)|<A\,\varepsilon^2, && j>r,\\
&A'<|\partial_\varepsilon(\lambda^{[n]}_j(x;\varepsilon)-\lambda^{[n]}_i(x;\varepsilon))|, && i\ne j>r,\\
&|\lambda^{[n]}_j(x;\varepsilon)-\lambda^{[n-1]}_j(x;\varepsilon)|\le\varepsilon^2 B\,e^{-\kappa_1 2^{n/\tau}}, && j>r,\\
&|\lambda^{[n]}_j(x;\varepsilon)|<A\,\min\{\varepsilon^2,\varepsilon\,x^2a_s^{-1}\}, && j\le r,
\end{aligned} \tag{5.14}$$

where $A',A>0$ are n,n_0-independent constants, and $\tau=\tau_0$.

Remarks. (1) The first three bounds on the eigenvalues in (5.14), follow from the first line of (5.13) by using the self-adjointness of the matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$; see Appendix A4. The other bounds in (5.13) imply the last bound in (5.14); see Appendix A4.
(2) The natural domain of definition in x of $\mathcal{M}^{[n]}(x,\varepsilon)$, $n>0$, will turn out to be $D(x)\le 2^{-2(n-1)}C_0^2$, but we imagine that it is defined for all x by continuing it as a constant from its limit value. In fact this is not important because, as we shall see, only the values of $\mathcal{M}^{[n]}(x,\varepsilon)$ with $D(x)\le 2^{-2(n-1)}C_0^2$ enter into the analysis. Smoothness means differentiability in $\varepsilon\in I_C$ and a right and left differentiability in x. The lack of differentiability in x, but the existence of right and left x derivatives, is due to the fact that the function $D(x)$ admits right and left derivatives: hence lack of differentiability in x appears as an artifact of the method. This lack of smoothness (unpleasant but inessential for our purposes) can be eliminated by changing $D(x)$ into a new $\widetilde{D}(x)$ which is smooth for x^2 between successive $\underline{\lambda}_j(\varepsilon)$'s and, at the same time, it is bounded above and

[4] For instance if $r=s=2$ and $f(\boldsymbol{\alpha},\boldsymbol{\beta})=f_{\mathbf{0}}(\boldsymbol{\beta})+f_1(\boldsymbol{\beta})\cos\alpha_1+f_2(\boldsymbol{\beta})\cos\alpha_2$, *to lowest order in x,ε,* one has $M^{[\le n]}_{\alpha\alpha}(x;\varepsilon)=3\varepsilon^2x^2(2\omega_u^4)^{-1}[f_u^2(\boldsymbol{\beta})+|\partial_\beta f_u(\boldsymbol{\beta})|^2]\delta_{u,v}$, $M^{[\le n]}_{\alpha\beta}=i\varepsilon^2x(2\omega_v^3)^{-1}\partial_{\beta_v}[(f_u^2(\boldsymbol{\beta})+|\partial_\beta\varphi_u(\boldsymbol{\beta})|^2)]$, and $M^{[\le n]}_{\beta\beta}=\varepsilon\partial^2_\beta f_0(\boldsymbol{\beta})$, $u,v=1,2$.

below proportionally to $D(x)$. But this would make the discussion needlessly notationally involved and we avoid it.
(3) One should also remark that, although we excluded some values of ε (i.e. we required $\varepsilon \in \mathcal{E}_{\overline{n}_0-1}$, see (3.3)), here all $\varepsilon \in I_C$ are allowed. The restriction on ε plays no role in the high frequency resummations: so far its only purpose is to avoid divisions by 0 and to assign a finite value to contributions to $\mathbf{h}$ from trees with with propagators on scale $[\ge \overline{n}_0]$ (which could be infinite because of the lack of an infrared cut-off in their expressions; see the third of (5.7)).
(4) The bounds on the entries of $\mathcal{M}^{[n]}(x;\varepsilon)$ in the second and third lines of (5.13) arise from cancellations that are checked in Ref. [GG] via a sequence of algebraic identities on the Lindstedt series coefficients and *the real difficulty lies in the proof of convergence*. The algebraic mechanism for the cancellations is briefly recalled in Appendix A3, for completeness.
(5) Loosely speaking (as mentioned in Section 4) the reason why the above result holds with $\overline{n}_0$-independent constants, and why its proof can be taken from Ref. [GG], is that if the scales of the propagators are constrained to be $[p]$ with $p < \overline{n}_0$ the propagators denominators can be estimated by $2^{-2(\overline{n}+1)-2}x^2$ by (4.4) and by the Remark (1) after Definition 1. This means that one can proceed as in the hyperbolic tori cases in which boundedness, from below, proportionally to x^2 of the propagators denominators was the main feature exploited and *no restriction* on ε had to be required, other than suitable smallness.

The lemma can be proved by imitating the convergence proof of the KAM theorem, see for instance Ref. [GG]; however in the following Appendix A3 the part of the proof which is not reducible to a purely algebraic check is repeated, for completeness.

We have therefore constructed a new representation of the formal series for the function $\mathbf{h}$ of the parametric equations for the invariant torus: in it only trees with lines carrying a scale label $[\infty],[0],\ldots,[\overline{n}_0-1]$ or $[\ge \overline{n}_0]$ and *no self-energy clusters* are present. The above lemma will be the starting block of the construction that follows.

6. Renormalization: the infrared resummation

Convergence problems still arise from the propagators $g^{[\ge \overline{n}_0]}(x;\varepsilon)$, which become uncontrollably large for $x = \boldsymbol{\omega}\cdot\boldsymbol{\nu}$ close to the eigenvalues of M_0 because the bound (4.4) which allowed control of the divisors in terms of the classical small divisors (i.e. in terms of $|x|$) does not hold any more. Hence we must change strategy.

DEFINITION 6. *Given $d\times d$ Hermitian matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$, $n=\overline{n}_0,\overline{n}_0+1,\ldots$, with eigenvalues $\lambda_j^{[n]}(x;\varepsilon)$, we introduce the following notations.*
(i) The sequence of self-energies $\underline{\lambda}_j^{[n]}(\varepsilon)$ *is defined for $n\ge\overline{n}_0$ by*

$$\underline{\lambda}_j^{[n]}(\varepsilon) \stackrel{def}{=} \lambda_j^{[n]}\Big(\sqrt{\underline{\lambda}_j^{[n-1]}(\varepsilon)},\varepsilon\Big), \qquad \underline{\lambda}_j^{[\overline{n}_0-1]}(\varepsilon) \stackrel{def}{=} \lambda_j^{[0]}, \tag{6.1}$$

provided $\underline{\lambda}_j^{[n]}(\varepsilon)\ge 0$, $n=\overline{n}_0,\overline{n}_0+1,\ldots$.
(ii) The propagator divisors *are defined for $n\ge\overline{n}_0$ by*

$$\Delta^{[n]}(x;\varepsilon) \stackrel{def}{=} \Big|x^2-\underline{\lambda}_{j(x)}^{[n]}(\varepsilon)\Big|, \tag{6.2}$$

where $j(x)$ is the label where the minimum of $\left|x^2-\underline{\lambda}_j^{[n]}(\varepsilon)\right|$ *is reached.*

Remarks. (1) The self-energies are defined recursively starting from those of the matrix M_0 whose first r eigenvalues are 0. Hence, *as long as one can extend the last of* (5.14) and as long as *the self-energies* $\underline{\lambda}_j^{[n]}(\varepsilon)$ *remain close to the original value* $\lambda_j^{[0]}$, as we shall check for ε small enough, one has $\underline{\lambda}_j^{[n]}(\varepsilon)=0$ for $j=1,\ldots,r$ and $\underline{\lambda}_j^{[n]}(\varepsilon)>0$ for $j>r$.
(2) Under the same conditions and if $\Delta^{n]}(x;\varepsilon)\simeq 2^{-2n}C_0^2$ the label $j(x)$ depends only on M_0, hence it is n-independent, and furthermore it is constant at x fixed, as ε varies in the intervals I introduced in Definition 1 (because for large n the frequency x is constrained to be close to one of the $\underline{\lambda}_j^{[n]}(\varepsilon)$).
(3) The name of *propagator divisor* assigned to $\Delta^{[n]}(x,\varepsilon)$ in (6.2) reflects its later use as a lower bound on the denominator of a propagator, see Remark (7) to the inductive assumption below.

By repeating the analysis of Section 4 we can represent the function $\mathbf{h}$ via sums of values of trees in which lines can carry scale labels $[\infty],[0],\ldots,[\overline{n}_0-1],[\overline{n}_0],[\overline{n}_0+1],\ldots$ and which contain no self-energy clusters and no trivial nodes (i.e. are renormalized trees, see Definition 5 in Section 5). The new propagators will be defined by the same procedure used to eliminate the self-energy clusters of scales $[n]$ with $n\le\overline{n}_0-1$. However we shall now determine the scale of a line in terms of the recursively defined $\Delta^{[n]}(x;\varepsilon)$ rather than in terms of $D(x)$: the latter becomes too rough to resolve the separation between the eigenvalues and their variations.

Let $X_{\overline{n}_0-1}(x)\stackrel{def}{=}\prod_{m=0}^{\overline{n}_0-1}\chi_m(D(x))$, $Y_n(x;\varepsilon)\stackrel{def}{=}\prod_{m=\overline{n}_0}^{n}\chi_m(\Delta^{[m]}(x;\varepsilon))$ for $n\ge\overline{n}_0$ and $Y_{\overline{n}_0-1}\equiv 1$: the definition of the new propagators will be

$$\begin{aligned}
g^{[\overline{n}_0]} &\stackrel{def}{=} X_{\overline{n}_0-1}(x)\,\psi_{\overline{n}_0}(\Delta^{[\overline{n}_0]}(x;\varepsilon))\,(x^2-\mathcal{M}^{[\le\overline{n}_0]}(x;\varepsilon))^{-1},\\
g^{[\overline{n}_0+1]} &\stackrel{def}{=} X_{\overline{n}_0-1}(x)\,\chi_{\overline{n}_0}(\Delta^{[\overline{n}_0]}(x;\varepsilon))\,\psi_{\overline{n}_0+1}(\Delta^{[\overline{n}_0+1]}(x;\varepsilon))\,(x^2-\mathcal{M}^{[\le\overline{n}_0+1]}(x;\varepsilon))^{-1},\\
&\ldots\\
g^{[n]} &\stackrel{def}{=} X_{\overline{n}_0-1}(x)\,Y_{n-1}(x;\varepsilon)\,\psi_n(\Delta^{[n]}(x;\varepsilon))\,(x^2-\mathcal{M}^{[\le n]}(x;\varepsilon))^{-1},
\end{aligned}\tag{6.3}$$

and so on, using indefinitely the identity $1\equiv\psi_n(\Delta^{[n]}(x;\varepsilon))+\chi_n(\Delta^{[n]}(x;\varepsilon))$ to generate the new propagators.

In this way we obtain a formal representation of $\mathbf{h}$ as a sum of tree values in which only renormalized trees (i.e. with neither trivial nodes nor self-energy lines, see Definition 4 in Section 4) and in which each line ℓ carries a *scale label* $[n_\ell]$. This means that we can formally write $\mathbf{h}$ as in (5.10), with $\mathrm{Val}(\theta)$ defined according to (5.11), but now the scale label $[n_\ell]$ is such that n_ℓ can assume all integer values ≥ 0 or ∞, and no line carries a scale label like $[\ge n]$: *only scale labels like* $[n]$ *are possible.* The corresponding propagators $g^{[n_\ell]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu}_\ell;\varepsilon)$ will be defined as follows.

Definition 7. *Given a sequence $\mathcal{M}^{[\le m]}(x;\varepsilon)$ as in Definition 6, $m\ge 1$, let $\mathcal{M}^{[n]}(x;\varepsilon)=\mathcal{M}^{[\le n]}(x;\varepsilon)-\mathcal{M}^{[\le n-1]}(x;\varepsilon)$ with $\mathcal{M}^{[\le 0]}\equiv\mathcal{M}^{[0]}\equiv M_0$ so that $\mathcal{M}^{[\le n]}(x;\varepsilon)=\sum_{m=0}^{n}\mathcal{M}^{[m]}\,(x;\varepsilon)$. Setting $\Delta^{[n]}(x;\varepsilon)\equiv D(x)$ if $n\le\overline{n}_0$, define for all $n\ge 0$*

$$g^{[n]}(x;\varepsilon)=\frac{\psi_n(\Delta^{[n]}(x;\varepsilon))\prod_{m\ge 0}^{n-1}\chi_m(\Delta^{[m]}(x;\varepsilon))}{x^2-\mathcal{M}^{[\le n]}(x;\varepsilon)}.\tag{6.4}$$

(for $n=0$ this means $\psi_0(D(x))\,(x^2-M_0)^{-1}$). We say that $g^{[n]}_\ell = g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu}_\ell;\varepsilon)$ is a propagator with scale $[n]$. The matrices $\mathcal{M}^{[m]}(x;\varepsilon)$ will be defined as in Section 5 for $n\le\overline{n}_0$ and will be defined recursively also for $n>\overline{n}_0$ in terms of the self-energy clusters $\mathcal{S}^{\mathcal{R}}_{k,n-1}$ introduced in Definition 4, Section 5, setting for $n>\overline{n}_0$ (see (5.9))

$$\mathcal{M}^{[n]}(x;\varepsilon) = \Big(\prod_{m=0}^{n}\chi_m(\Delta^{[m]}(x;\varepsilon))\Big)\sum_{k=2}^{\infty}\sum_{T\in\mathcal{S}^{\mathcal{R}}_{k,n-1}}\mathcal{V}_T(x;\varepsilon), \tag{6.5}$$

where the self-energy values $\mathcal{V}_T(x;\varepsilon)$ are evaluated by means of propagators on scales less than $[n]$. Note that we have already defined (consistently with (6.5))) the matrices $\mathcal{M}^{[\le n]}$ with $n\le\overline{n}_0$ and the propagators on scale $[\infty],[0],\ldots,[\overline{n}_0-1]$ (so that (6.4) defines also $g^{[\overline{n}_0]}(x;\varepsilon)$).

Remark. (1) Some propagators may vanish being proportional to a product of cut-off functions. If a propagator of a line has scale $[n]$ and does not vanish then, see (6.4),

$$2^{-2(n+1)}C_0^2 \le \Delta^{[n]}(x;\varepsilon) \le 2^{-2(n-1)}C_0^2 \tag{6.6}$$

(2) Our definitions of the matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$ for $n>\overline{n}_0$ will be such that given the node harmonics of a tree hence by (6.6) the scale $[n]$ that is attributed to a line can only assume up to two consecutive values unless the propagator (hence the tree value) vanishes.
(3) We may and shall imagine that scale labels are assigned arbitrarily to each line of a given tree *with the constraint that no self energy clusters are generated*; however the tree will have a non-zero value only if the scale labels are such that all propagators do not vanish. This means that only up to two consecutive scale labels can be assigned to each line if the tree value is not zero. The "ambiguity" on the scale labels for a line is due to the use of the non-sharp χ and ψ functions of Fig. 2.

We make an inductive assumption on the propagators on the scales $[m]$, $0\le m<n$.

Inductive assumption. *Let $\overline{n}_0\equiv n_0+\overline{n}$ (see (4.2)) and suppose n_0 large enough; then*
(i) For $0\le m\le n-1$ the matrices $\mathcal{M}^{[m]}(x;\varepsilon)$ are defined by convergent series for all $\varepsilon\in I_C$ and, for all x, they are Hermitian, and $\mathcal{M}^{[m]}(x;\varepsilon)=(\mathcal{M}^{[m]}(-x,\varepsilon))^T$. Furthermore they satisfy the same relations as (5.13), hence (5.14), with n replaced by m, for all $0<m<n-1$, with suitably chosen (new, possibly different) constants κ_1,A,A',B,τ. One can take $\tau=2\tau_1$.
(ii) There exist $K>0$ and open sets $\mathcal{E}^o_m$, $m=0,\ldots,n$, with $\mathcal{E}^o_m\subset I_C$, such that, defining recursively $\underline{\lambda}^{[m]}_j(\varepsilon)$ in terms of $\underline{\lambda}^{[m-1]}_j(\varepsilon)$ for $m=\overline{n}_0,\ldots,n-1$ by (i) in Definition 6 above, while setting $\underline{\lambda}^{[m]}_j(\varepsilon)\equiv\lambda^{[0]}_j$ for $m=0,\ldots,\overline{n}_0-1$, and defining $\tau_1\stackrel{def}{=}\tau_0+r+1$, see (3.4), one has for $\varepsilon\notin\mathcal{E}^o_m$

$$\begin{aligned}&\Gamma^{[m]}(x;\varepsilon)=\min\Big\{\min_j\Big|x\pm\sqrt{\underline{\lambda}^{[m]}_j(\varepsilon)}\Big|,\ \min_{j\ge i}\Big|x\pm\sqrt{\underline{\lambda}^{[m]}_j(\varepsilon)}\pm\sqrt{\underline{\lambda}^{[m]}_i(\varepsilon)}\Big|\Big\}\ge 2^{-\frac12 m}\frac{C_0}{|\boldsymbol{\nu}|^{\tau_1}},\\ &\qquad|\mathcal{E}^o_m|\le K2^{-\frac12 m}C^2,\end{aligned} \tag{6.7}$$

for all $m\le n-1$ and all x.

Remarks. Assuming validity of the hypothesis for $m<n$ we note a few of its implications.
(1) So far we have only checked the hypothesis for scales $[m]$ with $m\le\overline{n}_0$, as expressed by Lemma

2 in Section 5, i.e. for the high frequency propagators. If (i) is proved also for $m = n$ then we can impose (6.7) immediately by excluding a set $\mathcal{E}^o_m$ of ε's of measure estimated by $2^{-\frac{1}{2}m}C^2K$ with K a constant that can be bounded in terms of A', A by introducing the constants ρ_m and ρ'_m as in (A2.1), with $\underline{\lambda}^{[0]}_j(\varepsilon)$ replaced by $\underline{\lambda}^{[m]}_j(\varepsilon)$, and proceeding as done in Appendix A2 for the case $n \le \overline{n}_0$. Note that since the self-energies $\underline{\lambda}^{[m]}_j(\varepsilon)$ are $\equiv \underline{\lambda}^{[0]}_j(\varepsilon)$ for all $m = 0, \ldots, \overline{n}_0 - 1$ one will have, for such m's, $\mathcal{E}^o_m \equiv I_C/\mathcal{E}_{\overline{n}_0-1}$, see (3.3). It is very important to keep in mind, in the above argument, that the self-energies either are 0 (for $j \le r$) or are close within $O(\varepsilon^2)$ to the positive eigenvalues of M_0, and they are *differentiable* in ε and to the right and left of each x by (i); see (5.14).

(2) If a line with a scale $[n]$ has vanishing propagator (i.e. $g^{[n]}(x;\varepsilon) = 0$ because of the χ, ψ cut-off functions in the definition (6.4)) but n differs at most by one unit from the integer n' such that $g^{[n']}(x;\varepsilon) \ne 0$. Thus if we consider $\Delta^{[n]}(x,\varepsilon)$ we can bound it by changing the inequalities (6.6) into $C_0^2\, 2^{-2(n+2)} < \Delta^{[n]}(x;\varepsilon) \le C_0^2\, 2^{-2(n-2)}$. The remark will be useful later when we shall exploit it in the discussion of the cancellations which we shall study to check the inductive hypothesis.

(3) By (5.13) and (5.14), and (I) in Appendix A4, we deduce that $\lambda^{[m]}_j(x;\varepsilon)$, hence $\underline{\lambda}^{[m]}_j(\varepsilon)$, do not change by more than $C\,B\,\varepsilon^2 \sum_{n \ge \overline{n}_0} e^{-\kappa_1 2^{n/(2\tau_1)}}$, with respect to $\lambda^{[0]}_j(\varepsilon)$, if $\varepsilon < \overline{\varepsilon}_1$ (and $m \ge \overline{n}_0$).

(4) Hence if ε is small enough the self-energies, i.e. $\underline{\lambda}^{[m]}_j(\varepsilon)$, have distance bounded above by $2a_s\varepsilon$ and below by $\frac{1}{2}\varepsilon \min\left\{a_1, \min_j\{a_{j+1} - a_1\}\right\} = 2\rho\,\varepsilon a_s$ with ρ defined in (4.2), if ε is small enough, say $\varepsilon < \overline{\varepsilon}_2$.

(5) Therefore by Remark (4) we see that the distance of $|x|^2$ from the closest value $\underline{\lambda}^{[m]}_j(\varepsilon)$ is smaller than one fourth, up to corrections $O(\varepsilon^2)$, the distance between the distinct values of $\underline{\lambda}^{[m]}_j(\varepsilon)$, if m is large enough compared to n_0, i.e. if $2C_0^2 2^{-2m} < \rho\varepsilon a_s$ (or $m - n_0 \ge \overline{n}$ as implied by the definition (4.2) of $\overline{n}$). This means that $j(x)$ is m,ε-independent and it coincides with the label minimizing $\left|x^2 - |\lambda^{[m]}_j(x;\varepsilon)|\right|$ for all $m \ge \overline{n}_0$ and all $\varepsilon \in I$.

(6) $\underline{\lambda}^{[\overline{n}_0-1]}_j(\varepsilon) \equiv \lambda^{[0]}_j$ are x-independent and, by their definition, the same remains true for *all* $\underline{\lambda}^{[m]}_j(\varepsilon)$. The self-energy $\underline{\lambda}^{[m]}_j(\varepsilon)$ will be thought of as a reference position for the j-th eigenvalue on scale $[m]$, $m \le n-1$.

(7) As noted in Remark (5) the quantity $\left|x^2 - \lambda^{[n]}_{j(x)}(x;\varepsilon)\right|$ is the smallest denominator appearing in the value of the propagator of a line with momentum $\boldsymbol{\nu}$ if $g^{[n]}(x;\varepsilon) \ne 0$ (here $x = \boldsymbol{\omega}\cdot\boldsymbol{\nu}$). *The key to the analysis is the check that the quantities* $\Delta^{[n]}(x;\varepsilon)$ *can be used to bound below the denominators of the non-vanishing propagators of scale* $[n]$. If $\lambda^{[n]}_{j(x)}(x;\varepsilon) < 0$ one has $x^2 - \lambda^{[n]}_{j(x)}(x;\varepsilon) \ge x^2$, so that the assertion is trivially satisfied: therefore the really interesting case is when $\lambda^{[n]}_{j(x)}(x;\varepsilon) \ge 0$ (which includes the cases $j(x) > r$). If x has scale $[n]$ with $n \ge \overline{n}_0$ one has

$$
\begin{aligned}
\left| |x| - \sqrt{|\lambda^{[n]}_{j(x)}(x;\varepsilon)|} \right| &\ge \left| |x| - \sqrt{\underline{\lambda}^{[n]}_{j(x)}(\varepsilon)} \right| - \left| \sqrt{\underline{\lambda}^{[n]}_{j(x)}(\varepsilon)} - \sqrt{\lambda^{[n]}_{j(x)}(x;\varepsilon)} \right| \\
&\ge \frac{1}{2}\left| |x| - \sqrt{\underline{\lambda}^{[n]}_{j(x)}(\varepsilon)} \right| + 2^{-(n+3)}C_0 - \left| \sqrt{\lambda^{[n]}_{j(x)}\big(\sqrt{\underline{\lambda}^{[n-1]}_{j(x)}(\varepsilon)}, \varepsilon\big)} - \sqrt{\lambda^{[n]}_{j(x)}(x;\varepsilon)} \right| \\
&\ge \frac{1}{2}\left| |x| - \sqrt{\underline{\lambda}^{[n]}_{j(x)}(\varepsilon)} \right|, \qquad \Rightarrow \qquad \|x^2 - \mathcal{M}^{[n]}(x,\varepsilon)\| \ge \frac{1}{2^3}\sqrt{\frac{a_1}{a_s}}\left|x^2 - \underline{\lambda}^{[n]}_{j(x)}(\varepsilon)\right|,
\end{aligned}
\tag{6.8}
$$

having used the lower cut-off $\psi_n(\Delta^{[n]}(x;\varepsilon))$ in the propagator (see (6.3)) to obtain the first two terms in the second line (and added a further factor 2^{-1} in order to extend the result also to the propagators considered in Remark (3)), while the upper cut-off $\chi_{n-1}(\Delta^{[n-1]}(x;\varepsilon))$ has been used

to obtain positivity of the difference between the second and third terms in the second line, after applying (5.14), for $n \ge \overline{n}_0$, to get

$$\max_x |\partial_x^{\pm} \lambda_{j(x)}^{[n]}(x;\varepsilon)| \le \widehat{B}\,\varepsilon^2,\ j(x) > r, \qquad |\lambda_{j(x)}^{[n]}(x;\varepsilon)| \le \widehat{B}\,\varepsilon\,|x|^2 \ \le \varepsilon\, C_0^2\, 2^{-2n},\ j(x) \le r, \tag{6.9}$$

for some $\widehat{B}$, so that the last term in the second line of (6.8) can be bounded above proportionally to $\varepsilon 2^{-n} C_0$. Hence the first inequality in the last line of (6.8) follows if ε small enough, say $\varepsilon \le \overline{\varepsilon}_3$ for some $\overline{\varepsilon}_3$, *fixed independently of* n. The latter constraint can be achieved simply by taking n_0 large enough, see (3.2). The last implication follows from (6.9) if $j(x) \le r$ because $\underline{\lambda}_{j(x)}^{[n]}(\varepsilon) = 0$. Otherwise if $j(x) > r$ and $|x|, \sqrt{\underline{\lambda}_{j(x)}^{[n]}(\varepsilon)}, \sqrt{\underline{\lambda}_{j(x)}^{[n]}(x,\varepsilon)} \in [\frac{1}{2}\sqrt{\varepsilon\, a_1}, 2\sqrt{a_s\,\varepsilon}]$ one has $(|x| + \sqrt{\underline{\lambda}_{j(x)}^{[n]}(x,\varepsilon)})/(|x| + \sqrt{\underline{\lambda}_{j(x)}^{[n]}(\varepsilon)}) \ge 2^{-2}\sqrt{a_1/a_s}$, as long as $\varepsilon < \overline{\varepsilon}_2$ (see Remark (4) above): implying again the (6.8). Hence $\Delta^{[n]}(x;\varepsilon)$ can be effectively used to estimate the size of the non-vanishing propagators which is, therefore, closely related to the scale of the corresponding lines.
(8) The Diophantine condition (3.3) and (6.7) will play from now on a key role. We begin by remarking that if the inductive hypothesis is proved *all lines will eventually acquire a well defined scale label*: in fact fixed x one cannot have $\Delta^{[n]}(x,\varepsilon) \le 2^{-2n} C_0^2$ for all n because this implies[5] $||x| - \sqrt{\underline{\lambda}_{j(x)}^{[n]}(\varepsilon)}| < 2^{-n} C_0$, which soon or later becomes incompatible with the first of (6.7). This explains why there is no trace left of the propagators $g^{[\ge n]}(x,\varepsilon)$.

To estimate the corrections to the self-energy as n increases it is clear that we must estimate the size of $\mathcal{M}^{[n]}(x;\varepsilon)$. For this purpose we need the following result.

LEMMA 3. *There is $\overline{\varepsilon}$ small and constants κ_1, A, A', B such that if $\varepsilon < \overline{\varepsilon}$ and the inductive hypothesis is assumed for $0 \le m \le n-1$ then the matrix $\mathcal{M}^{[n]}(x;\varepsilon)$ can be bounded by (5.13) and the inductive hypothesis holds for $m = n$.*

Hence the hypothesis holds for all n since we have already checked it for $n = 0, \ldots, \overline{n}_0$ (Lemma 2): the new constants κ_1, A, A', B will be different from the ones determined in Lemma 2.

Proof. For $n \le \overline{n}_0$ the bound (5.13) is covered by Lemma 2. So we can assume $n \ge \overline{n}_0 + 1$. Suppose first $\varepsilon \in \cap_{m=\overline{n}_0-1}^{n-1} \mathcal{E}_m$, with $\mathcal{E}_m = I_C \setminus \mathcal{E}_m^o$, so that the Diophantine property (6.7) holds for all $m \le n-1$. Consider a self-energy cluster T in $\cup_{k=2}^{\infty} \mathcal{S}_{k,n-1}^{\mathcal{R}}$. If the entering and exiting lines (with propagators of scale $[\ge n]$) have momenta $\boldsymbol{\nu}$ we begin by showing that

$$\sum_{\mathfrak{v} \in V(T)} |\boldsymbol{\nu}_{\mathfrak{v}}| > 2^{(n-6)/(2\tau_1)}. \tag{6.10}$$

Indeed the cluster contains at least one line $\ell = \ell_{\mathfrak{v}}$ with propagator which we can suppose to be not vanishing and which has scale $[n-1]$. We can write $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell^0 + \sigma_\ell \boldsymbol{\nu}$, where $\sigma_\ell = 0, 1$ and we set $\boldsymbol{\omega} \cdot \boldsymbol{\nu} = x$, $\boldsymbol{\nu}_\ell^0 = \sum_{\substack{\mathfrak{w} \in V(T) \\ \mathfrak{w} \preceq \mathfrak{v}}} \boldsymbol{\nu}_{\mathfrak{w}}$, and finally $x_\ell = \boldsymbol{\omega} \cdot \boldsymbol{\nu}_\ell$.

Since the line ℓ is not on scale $[n-2]$ (as it is on scale $[n-1]$) it follows from (6.3) that

$$\big||x_\ell| - \sqrt{\underline{\lambda}_{j(x_\ell)}^{[n-2]}(\varepsilon)}\big| \le 2^{-(n-2)} C_0, \tag{6.11}$$

[5] As $|a^2 - b^2| < c^2$ implies $|a - b| < c$ for $a, b, c > 0$.

Therefore if (6.10) does not hold and if $\sigma_\ell = 0$, by the first part of the Diophantine conditions (6.7), one finds $\big||x_\ell| - \sqrt{\underline{\lambda}_i^{[m]}(\varepsilon)}\big| > C_0 2^{-m/2} 2^{-(n-6)/2}$ for all $m \le n-1$ and for all $1 \le i \le d$, which would be in contradiction with (6.11).

If instead $\sigma_\ell = 1$ we shall use the second part of the Diophantine conditions (6.7) and get a contradiction. Remark that x can be assumed to be on scale $[q]$ with $q \ge n$ because of the cut-off functions in (6.5) so that one has $\big||x| - \sqrt{\underline{\lambda}_{j(x)}^{[p]}(\varepsilon)}\big| \le C_0 2^{-p}$ for $p \le n-1$. Hence if x_ℓ satisfies (6.11) we get, by assuming that (6.10) does not hold,

$$\begin{aligned} 2^{3-n} C_0 &\ge \Big| |x_\ell| - \sqrt{\underline{\lambda}_{j(x_\ell)}^{[n-2]}(\varepsilon)} \Big| + \Big| |x| - \sqrt{\underline{\lambda}_{j(x)}^{[n-2]}(\varepsilon)} \Big| \\ &\ge \Big| x_\ell - x + \eta_\ell \sqrt{\underline{\lambda}_{j(x_\ell)}^{[n-2]}(\varepsilon)} + \eta \sqrt{\underline{\lambda}_{j(x)}^{[n-2]}(\varepsilon)} \Big| \\ &\ge \frac{C_0}{2^{(n-2)/2} |\boldsymbol{\nu}_\ell - \boldsymbol{\nu}|^{\tau_1}} = \frac{C_0}{2^{(n-2)/2} |\boldsymbol{\nu}_\ell^0|^{\tau_1}} \ge 2^{4-n} C_0, \end{aligned} \tag{6.12}$$

for some $\eta, \eta_\ell = \pm 1$, which again leads to a contradiction, so that (6.10) holds also in such a case.

Every node factor contributes to $\mathcal{M}^{[n]}$ a factor $f_{\boldsymbol{\nu}_\mathfrak{v}}$ bounded by $F_0 e^{-\kappa_0 |\boldsymbol{\nu}_\mathfrak{v}|}$; there are $\le (4d^2)^k k!$ self-energy clusters, 4^k scales (for each line there are only two scales for which the propagator is not zero, and one has to allow also a scale different by one unit from that which corresponds to have a nonvanishing propagator, see Remark (3) after the inductive assumption), and $\mathcal{N}_m(T)$ lines of scale $m = 0, 1, \ldots, n$ in each self-energy cluster T contributing to $M^{[n]}(x;\varepsilon)$ and not to the $M^{[m]}(x;\varepsilon)$, with $m < n$. Thus the bound on $M^{[n]}(x;\varepsilon)$ is

$$G_0 \sum_{k=2}^{\infty} \varepsilon^k G_1^k e^{-\frac{1}{2}\kappa_0 \sum_{\mathfrak{v} \in V(T)} |\boldsymbol{\nu}_\mathfrak{v}|} e^{-G_2 2^{n/(2\tau_1)}} \prod_{m=0}^{n} 2^{2m \mathcal{N}_m(T)}, \tag{6.13}$$

for suitable constants G_0, G_1, G_2, explicitly computable by the above remarks; for $k < 4$ the exponent of ε can be replaced by 2, see Remark after Definition 4. The estimate of the number $\mathcal{N}_m(T)$ is given in Appendix A3 (cf. in particular Section A3.4), and gives $\mathcal{N}_m(T) \le E_m \sum_{\mathfrak{v} \in V(T)} |\boldsymbol{\nu}_v|$, with $E_m = 2^{(6-m)/(2\tau_1)}$ and τ_1 in (3.4), which shows convergence of the series in (6.13) if ε is small enough, say $\varepsilon < \overline{\varepsilon}$.

We can and shall assume that $\overline{\varepsilon}$ does not exceed $\min\{\overline{\varepsilon}_1, \overline{\varepsilon}_2, \overline{\varepsilon}_3\}$, with $\overline{\varepsilon}_1$, $\overline{\varepsilon}_2$ and $\overline{\varepsilon}_3$ introduced earlier (see Remarks 3, 4, 7 after the inductive hypothesis). *The rest of the argument repeats the analysis in Appendix A3 with minor notational changes*: we only hint at the details in Appendix A3.4.

Under the considered hypotheses the matrices $\mathcal{M}^{[n]}(x;\varepsilon)$ are well defined, by the above discussion on convergence of the defining series on the set $\cap_{m=\overline{n}_0-1}^{n-1} \mathcal{E}_m$. The symmetry in item (i) is due to algebraic identities valid for the Lindstedt series. They are detailed in Ref. [GG], Appendix A5, for $\varepsilon < 0$: being of algebraic nature the argument does not depend on the sign of ε and it holds unchanged in the present case.

The second and third lines of inequalities in (5.13) embody the cancellations. We need to check the cancellations to make sure for instance that the structure of the matrix $\mathcal{M}^{[n]}(x;\varepsilon)$ preserves the eigenvalues and the Whitney smoothness: a danger being that the first r eigenvalues become "detached" from 0, i.e. no longer can be bounded by εx^2, see (5.14). For instance a bound like $O(\varepsilon^2)$ would not be enough as it would imply that the self-energies $\underline{\lambda}_j^{[n]}(\varepsilon)$ may become different from zero for $j \le r$.

Since the function $\mathcal{M}^{[n]}(x;\varepsilon)$ is defined on the complement of a dense open set differentiability in the sense of Whitney can be proved (as usual) by computing a formal derivative and then showing that it is continuous and that it can also be used as a bound in interpolations.[6]

The computation of the formal derivatives proceeds as the computation of the actual derivatives done in the proof of Lemma 2 (in Appendix A3). One proves formal right and left continuous differentiability of the matrices $\mathcal{M}^{[n]}(x;\varepsilon)$ on the closed set $\cap_{m=\overline{n}_0-1}^{n-1}\mathcal{E}_m$ simply by differentiating term by term the value of each cluster contributing to $\mathcal{M}^{[n]}(x;\varepsilon)$. This involves differentiating matrices like $(x^2-\mathcal{M}^{[\le p]}(x;\varepsilon))^{-1}$, i.e. the matrices $\mathcal{M}^{[p]}(x;\varepsilon)$ with $p<n$, which are differentiable by the inductive assumption, or it involves differentiating the cut-off functions ψ_p,χ_p and the locations $\underline{\lambda}_j^{[p]}(\varepsilon)$ with $j>r$ (because $\underline{\lambda}_j^{[p]}(\varepsilon)\equiv 0$ for $j\le r$) which appear in the form $\Delta^{[p]}(x,\varepsilon)$ in the arguments of the cut-off functions. All such quantities are differentiable in ε and right and left differentiable in x by the inductive assumption; furthermore all terms arising from differentiation either of $\mathcal{M}^{[p]}(x;\varepsilon)$ or $\underline{\lambda}_j^{[p]}(\varepsilon)$, with $p<n$, appear multiplied by some power of ε, so that the inductive assumption is found to hold also for $p=n$ (for a similar discussion see Ref. [Ge]).

Note that $\Delta^{[n]}(x;\varepsilon)$ depend on $j(x)$ but as ε varies within the interval I, see (ii) in definition 1, $j(x)$ is not only ε-independent but it is also constant in x for x varying in small intervals near the eigenvalues of M_0 and, therefore, in intervals widely spaced because $n\ge\overline{n}_0$: this is due to the cut-off functions which force x to be close to a single eigenvalue if the propagator of the corresponding line is different from 0. Hence for $n\ge\overline{n}_0$ we do not have to differentiate the function $j(x)$ (neither with respect to x nor with respect to ε from which it does not depend); for $n<\overline{n}_0$ the function $j(x)$ is constant to the right and to the left of every point.

The n-independence of the constants A',A,B appearing in the inductive hypothesis is proved word by word as the corresponding argument in Appendix A3; the constant κ_1 has been estimated above (see G_2 in (6.13)) and is n-independent.

The interpolation bound, see footnote [6], necessary for defining the Withney derivatives, holds because in comparing two contributions to $\mathcal{M}^{[n]}(x;\varepsilon)$ with different x or different ε the difficulty might only come from the comparison of $(x_\ell^2-\mathcal{M}^{[\le p]}(x_\ell,\varepsilon))^{-1}$ evaluated at two different points and for one line ℓ at a time: this can be done algebraically by using the resolvent identity

$$\begin{aligned}\Big(x_\ell^2-\mathcal{M}^{[\le p]}(x_\ell,\varepsilon)\Big)^{-1}-\Big({x'}_\ell^2-\mathcal{M}^{[\le p]}(x'_\ell,\varepsilon')\Big)^{-1}&=\Big(x_\ell^2-\mathcal{M}^{[\le p]}(x_\ell,\varepsilon)\Big)^{-1}\cdot\\ \cdot\Big({x'}_\ell^2-x_\ell^2+\mathcal{M}^{[\le p]}(x'_\ell,\varepsilon')-\mathcal{M}^{[\le p]}(x_\ell,\varepsilon)\Big)&\Big({x'}_\ell^2-\mathcal{M}^{[\le p]}(x'_\ell,\varepsilon')\Big)^{-1},\end{aligned}\tag{6.14}$$

which involves only denominators evaluated at x,ε's which are in the set where they are controlled by the (6.7) and therefore can be estimated in the same way as the formal derivatives. The Withney extension is therefore possible keeping control of the bounds for all ε's (small as above) and x. The dependence on x may involve the functions $D(x)$ (for $p\le\overline{n}_0-1$) so that the differentiability in x will be possible only to the right and to the left of each point (this involves a natural generalization of Whitney's theorem).

The cancellations analysis (i.e. the proof of the second and third inequalities in (5.13)) is inductive and has been performed several times in the literature, see Refs. [Ga] and [GG]: in Appendix A3

[6] More precisely in its simplest form Whitney's theorem states that if $F(x)$ is a function defined on a closed set C of the interval $[0,1]$ and if there is a continuous function $F'(x)$ defined on C and if for some $\gamma>0$ and all $x,y\in C$ one has $|F(y)-F'(x)(y-x)|<\gamma|x-y|$ (we call this an *interpolation bound*) then there is a continously differentiable function $\overline{F}(x)$ extending F to $[0,1]$ and with derivative $\overline{F}'(x)$, with $\max|\overline{F}'(x)|<\gamma$, extending $F'(x)$.

we have repeated it following the version in Ref. [GM1] with some minor modifications. The same proof applies to the present case (being a purely algebraic check).

The inequalities (5.13) imply the (5.14) and therefore we get differentiability of the matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$ and of the self-energies. This allows us to impose validity of (6.7) by excluding a few more values of ε by Remark (1) to the inductive hypothesis.

Therefore we conclude that $\mathcal{M}^{[n]}(x;\varepsilon)$ is defined and verifies (5.13) (with suitably chosen constants κ_1, A', A, B) in the same domain $\varepsilon < \overline{\varepsilon}$, where the matrix $\mathcal{M}^{[\le p]}(x;\varepsilon)$ is already defined for $p \le n-1$. *Of course* $\mathcal{M}^{[n]}$ *will be relevant for our analysis only on the set* $\cap_{m=\overline{n}_0-1}^{n}\mathcal{E}_m$ and the extension outside such set is only useful to simplify the analysis as it allows us to use freely interpolations formulae, mainly to check (6.7). The matrix $\mathcal{M}^{[\le n-1]}(x;\varepsilon)$ verifies the inductive assumption although it has physical meaning only for $\varepsilon \in \cap_{m=\overline{n}_0-1}^{\infty}\mathcal{E}_n$, where $\mathcal{E}_n$ is the domain in which (6.7) holds for $m \le n$. ■

Having checked that the series defining the $\mathcal{M}^{[\le n]}(x;\varepsilon)$, hence the self-energies, converge and verify the bounds in the inductive hypothesis we still have to check that the fully renormalized series for $\mathbf{h}$, which has thus been shown to make sense term by term, converges and that its sum is actually a function $\mathbf{h}$ satisfying the equations for the parametric representation of invariant tori.

To study convergence we can take again advantage of the method, already used in the proof of Lemmas 2,3 above to estimate the number of lines on scale n in a self-energy cluster containing no self-energy clusters. Indeed also for renormalized trees one can prove a bound like $\mathcal{N}_m(\theta) \le E_m \sum_{\mathfrak{v}\in V(\theta)} |\boldsymbol{\nu}_v|$ for the number $\mathcal{N}_m(\theta)$ of lines on scales m contained in $\Lambda(\theta)$ with E_m fast decreasing with m: $E_m \stackrel{def}{=} 2^{(6-m)/(2\tau_1)}$ (see Appendix A3). Hence convergence in the region $\mathcal{E} \in \cap_{n=\overline{n}_0-1}^{\infty}\mathcal{E}_n$ follows because if we only sum values of trees without self-energy clusters then we can use the above bound on $\mathcal{N}_m(\theta)$.

The renormalized trees may contain lines of scales $[\infty]$ which so far played no role. Their propagators are bounded below by a constant times ε; however their number cannot be larger than $\frac{1}{2}k$ in trees of order k (see Remark (5) in Section 2). Therefore they may reduce the factor ε^k normally present in the value of a graph with k nodes to $\varepsilon^{\frac{1}{2}k}$; hence this will not affect the convergence of the series other than by putting a more severe constant on the maximum value of ε.

The set $\mathcal{E}_n^0$, complement of $\mathcal{E}_n$ in I_C, has measure estimated by $C^2 2^{-n/2} K$ for $\varepsilon \in [(\frac{1}{2}C)^2, C^2] = I_C$. Since $C = 2^{-n_0}C_0$ and $n \ge \overline{n}_0 - 1 > n_0$ this is a very small fraction of the interval I_C and the smaller the closer is I_C to 0. This means that the set of ε's for which the whole construction can be performed has 0 as a density point. Note that the resummation just defined is a real resummation of our series only for $\varepsilon \in \cap_{n=\overline{n}_0-1}^{\infty}\mathcal{E}_n$, and there it gives a well defined function.

The check that the functions $\mathbf{h}(\boldsymbol{\psi})$ defined by the convergent renormalized series evaluated at $\boldsymbol{\psi} = \boldsymbol{\omega} t$ do actually solve the equations of motion can be performed by repeating the corresponding analysis in Ref. [Ge]. The equation that $\mathbf{h} = (\mathbf{a}, \mathbf{b})$ has to solve is $\mathbf{h} = \varepsilon g(\partial_{\boldsymbol{\alpha}} f(\boldsymbol{\psi} + \mathbf{a}, \boldsymbol{\beta}_0 + \mathbf{b}), \partial_{\boldsymbol{\beta}} f(\boldsymbol{\psi}+\mathbf{a}, \boldsymbol{\beta}_0+\mathbf{b}))$ where g is the pseudo-differential operator $(\boldsymbol{\omega}\cdot\boldsymbol{\nu})^{-2}$. The proof is of algebraic nature and ultimately follows from the fact that the series we are considering is a resummation of Lindstedt's series which is a formal solution of the problem. This explains why the various algebraic identities necessary for the check actually hold and the proof proceeds exactly as in Section 8 of Ref. [Ge]: we reproduce the argument and the chain of identities in Appendix A5. Therefore the

proof of Theorem 1 in Section 1 is complete.

7. Concluding remarks

The analysis can be immediately extended to the case in which the matrix $\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta_0})$ has some non-degenerate positive eigenvalues and some additional negative ones. The negative eigenvalues give no problems and they can be treated as in the case of Ref. [GG] in which all eigenvalues are negative. The negative eigenvalues do not give rise to new small divisors, unlike the positive ones; in more physical language the proper time scales (i.e. real proper frequencies) of the tori cannot resonate with the time scales of hyperbolic type (i.e. imaginary) introduced by the perturbation. Hence the following generalization of Theorem 1 holds.

Theorem 2. *If the matrix $\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta}_0)$ is not singular and has pairwise distinct eigenvalues the conclusions (i), (ii) and (iii) of Theorem 1 in Section 1 follow also in this case.*

The present work has developed a combinatorial approach to the proof that the frequencies of elliptic type possibly introduced by the perturbation do not resonate with the proper frequencies of the tori at least if ε is not too special in a small interval $[0,\overline{\varepsilon}]$. i.e. if it is in a set $\mathcal{E}\subset[0,\overline{\varepsilon}]$ of large measure near 0: Nevertheless the complement of $\mathcal{E}$ is an open dense set in $[0,\overline{\varepsilon}]$. The results hold for the Hamiltonian (1.2) and the special resonances $(\boldsymbol{\omega},\mathbf{0})$ considered: they can be extended to the most general resonances of Hamiltonians like (1.2) with a general quadratic form for the kinetic part (i.e. with $\frac{1}{2}\mathbf{I}\cdot\mathbf{I}$ replaced by $\frac{1}{2}\mathbf{I}\cdot Q\mathbf{I}$ with Q a non-degenerate $d\times d$ matrix).

The case of $\partial^2_{\boldsymbol{\beta}} f_{\mathbf{0}}(\boldsymbol{\beta_0})$ with degenerate eigenvalues seems quite different from the one treated here. Degeneracy will be removed to order $O(\varepsilon^2)$ under generic conditions. However $O(\varepsilon^2)$ is also the order of variation of the self-energies and one has to find a way to perform the resummations even between scale $\overline{n}_0$ and scale $2\overline{n}_0$, which is the scale at which the singularities of the propagator are split apart and one shall be able to proceed in the same way as we did in the case of non-degenerate eigenvalues.

The Lipschitz regularity in ε in Theorems 1 and 2 can be replaced by C^k regularity for any k by exploiting the comments in Remark (2) to Lemma 2 and Remark (2) in Appendix A3.2.

Unfortunately there seems to be no example known in which one can check that the power series studied here are divergent *as power series* in ε. Note that the (*infinitely many*) divergent series that have arisen in this paper are obtained by first splitting the coefficient of order k in the Lindstedt power series and then collecting contributions from the different orders in ε: the latter form divergent series for which we have assigned a summation rule. Therefore we have not proved divergence of the Lindstedt series as power series in ε: in this sense (*unlikely*) convergence of the Lindstedt series has not been ruled out (*yet*). Nor there is any uniqueness result on the value of the renormalized series. The latter depends on quite a few arbitrary choices (even in the hyperbolic cases): for instance the cut-off shapes in Fig. 2 are quite arbitrary and in principle the allowed ε's will change with the choice.

Furthermore, although we have not really checked all necessary details, it seems to us that our method also shows that, given a value ε_0 for which the renormalized series converges, one can find a complex domain of ε which is open, reaches the real axis with a vertical cusp at ε_0 and extends to an open region including a segment $(-\eta,0)$ on the negative real axis. In this domain the renormalized

series should converge taking on the real axis real values parameterizing an *hyperbolic* torus with the same rotation vector. However since there are no uniqueness proofs we cannot guarantee that each such extension does not correspond to a *different* torus (close within any power of ε to any other torus of the same type). This would signal a "giant bifurcation" that one would like to exclude: in Ref. [GG] an attempt was made to show uniqueness by estimating the size of the Lindstedt series coefficients aiming at applying the theory of Borel transforms. However we could not prove good enough bounds. We obtained $k!^\alpha$ growth with a too large α (given our estimated size of the domain of analyticity in ε) to apply uniqueness results from the theory of Borel summations .

Appendix A1. A brief review of earlier results

The system which is usually studied in literature when the problem of persistence of lower-dimensional elliptic tori is studied, is of the form

$$\mathcal{H} = \boldsymbol{\omega}(\xi) \cdot \mathbf{A} + \sum_{k=1}^{s} \Omega_k(\xi) \left(q_k^2 + p_k^2 \right) + P(\boldsymbol{\alpha}, \mathbf{A}, \mathbf{q}, \mathbf{p}), \tag{A1.1}$$

where $(\boldsymbol{\alpha}, \mathbf{A}, \mathbf{p}, \mathbf{q}) \in \mathbb{T}^r \times \mathbb{R}^r \times \mathbb{R}^s \times \mathbb{R}^s$. The function P is analytic in its arguments, and ξ is a parameter in $\mathbb{R}^r$; the function P is a *perturbation*: this means that a rescaling of the actions could allow us to introduce a small parameter ε in front of the function P. The frequencies of the harmonic oscillators are called *normal frequencies*; the case $\Omega_k(\xi) = \Omega_k =$ constant (that is with the normal frequencies independent of ξ) is a particular case, and it is usually referred to as the "constant frequency case". Existence of invariant tori for the system (A1.1) was originally proved by Mel'nikov [Me1], [Me2], new proofs were produced by Eliasson [E1], Kuksin [Ku], and Pöschel [P1]. The case $s = 1$ is easier, and it was earlier solved by Moser [Mo]. Later proofs were given by Rüssmann, see for instance Ref. [R]. See also the very recent Ref. [LW].

For $P = 0$ the dimension of the tori is $r < d$ and the variables $(\mathbf{q}, \mathbf{p})$ move around stable equilibrium points, hence such tori are called *elliptic lower-dimensional tori*.

The conditions under which the quoted results are proved are, besides the usual Diophantine condition (1.3) on $\boldsymbol{\omega}$, two non-resonance conditions involving one and two normal frequencies (the so called first and second Mel'nikov conditions, originally introduced in Ref. [Me1]); in particular one has to impose that the normal frequencies are non-degenerate (i.e. they have to be all different from each other).

Recently proofs of existence of elliptic lower-dimensional tori were given by requesting only the first Mel'nikov conditions: this allows treating degenerate frequencies. The first result in this direction is due to Bourgain [B3], where the ideas introduced in Refs. [CrW] and [B1] to prove existence of periodic and quasi-periodic solutions in nearly integrable Hamiltonian partial differential equations were adapted to construct lower-dimensional tori in the finite-dimensional Hamiltonian systems (A1.1) corresponding to the case of constant normal frequencies. New proofs, extending the results also to the case of non-constant normal frequencies, are due to Xu and You [Y], [XY].

An extension of the results of existence of periodic and quasi-periodic solutions describing lower-dimensional invariant tori for infinite-dimensional PDE systems has been provided in a series of papers, which include Refs. [Ku], [Wa], [CrW], [KP], [P2], [B1], [B2], [B4], [GM2] and [GMP].

On the other hand the problem (1.2) has not been widely studied in literature. It corresponds to a degenerate case because in absence of perturbations the lower-dimensional tori are neither

elliptic nor hyperbolic: it is the perturbation itself which determines if the tori, when continuing to exist, become elliptic or hyperbolic (or of mixed type or parabolic).
(i) The case of hyperbolic tori is easier, and it was the first to be studied, by Treshchëv [T]. Recently the problem was reconsidered in Ref. [GG], where the analyticity domain of the invariant tori was studied in more detail. In the case of elliptic tori the problem was considered in Refs. [ChW] and [WC], where Treshchëv's approach to the study of the case of hyperbolic tori, involving a preliminary change of coordinates, is used to cast the Hamiltonian in a form which is suitable for applying Pöschel's results on elliptic tori: in particular this imposes the same conditions as in Ref. [P1] on the normal frequencies which appear after the canonical change of coordinates is performed.
(ii) The existence problem has been also considered in Ref. [JLZ], *where elliptic and hyperbolic tori were studied simultaneously*, again by imposing some non-degeneracy conditions on normal frequencies. Ref. [JLZ] does not investigate resummations of Lindstedt's series; it is based on a rapid convergence method, close in spirit to the original proofs of the KAM theorem: a concise existence proof of lower-dimensional tori is achieved in both the elliptic and hyperbolic cases.

We stress that in all quoted papers, except Ref. [JLZ] and [T], the problem is considered with ε (i.e. the size of the perturbation) fixed and the study deals with estimates of the measure of the rotation vectors $\boldsymbol{\omega}$ for which there exist invariant tori. We suppose, instead, that $\boldsymbol{\omega}$ is fixed, hence we study the dependence on ε of the lower-dimensional invariant tori and, in particular, the set of values of ε for which the tori survive.

Our techniques extend those in Refs. [GG] and [Ge], and are based on the method introduced in Refs. [E2] and [Ga]. With respect to Ref. [Ge], where existence of quasi-periodic solutions is proved for the *generalized Riccati equation* considered in Ref. [Ba], the main difficulty is due to the presence of several normal frequencies. It is not surprising that this generates extra technical difficulties: as already noted, it is well known that the case $s = 1$ is easier; see Refs. [Mo] and [C]. An advantage of the present method is that it is fully constructive and gives a very detailed knowledge of the solution.

Appendix A2. Excluded values of ε

Define

$$
\begin{aligned}
\rho_{\overline{n}_0-1} &\stackrel{def}{=} \sqrt{\frac{\varepsilon}{a_s}} \min\Big\{ \min_{i>r} |\partial_\varepsilon \sqrt{\underline{\lambda}_i^{[0]}(\varepsilon)}|, \min_{\substack{i\neq j\\ i,j>r}} |\partial_\varepsilon \sqrt{\underline{\lambda}_j^{[0]}(\varepsilon)} - \partial_\varepsilon \sqrt{\underline{\lambda}_i^{[0]}(\varepsilon)}| \Big\}, \\
\rho'_{\overline{n}_0-1} &\stackrel{def}{=} \frac{1}{\sqrt{\varepsilon a_s}} \max_j \Big\{ \sqrt{\underline{\lambda}_j^{[0]}(\varepsilon)} \Big\},
\end{aligned}
\tag{A2.1}
$$

and note that $\rho_{\overline{n}_0-1}$ is bounded from below proportionally to ρ, as defined in (4.2), and $\rho'_{\overline{n}_0-1} = 1$. Then (3.3) excludes, for each $\boldsymbol{\nu}$, an interval in ε whose measure is bounded (using $\sqrt{a_s\varepsilon} \le C$; see (3.2)) by

$$
2^{-(\overline{n}_0-1)/2} C\, C_0\, K_0 |\boldsymbol{\nu}|^{-\tau_1}, \tag{A2.2}
$$

where the constant K_0 can be estimated by $K_0 = s\, a_s^{-1} \rho_{\overline{n}_0-1}^{-1}$.

The Diophantine condition on $\boldsymbol{\omega}$ implies that if (3.3) is invalid then $|\boldsymbol{\nu}|$ cannot be too small

$$
2\sqrt{\varepsilon a_s \rho'_{\overline{n}_0-1}} + 2^{-(\overline{n}_0-1)/2} C_0 |\boldsymbol{\nu}|^{-\tau_1} \ge |x| \ge C_0 |\boldsymbol{\nu}|^{-\tau_0}. \tag{A2.3}
$$

Therefore $\sqrt{\varepsilon a_s \rho'_{\overline{n}_0-1}} \geq \frac{1}{4} C_0 |\boldsymbol{\nu}|^{-\tau_0}$ if $\overline{n}_0 \geq 3$, hence in this case we only have to consider the values of $\boldsymbol{\nu}$ with $|\boldsymbol{\nu}| \geq (C_0/(4\sqrt{\varepsilon a_s \rho'_{\overline{n}_0-1}}))^{1/\tau_0}$. Since $C/2 < \sqrt{\varepsilon a_s} \leq C = 2^{-n_0} C_0$, we get the bound (3.5) with $\tau_1 = \tau + r + 1$ and $K = K_0\, C_0\, (4C\sqrt{\rho'_{\overline{n}_0-1}}\, C_0^{-1})^{(\tau_1 - r - 1)/\tau_0} \sum_{\boldsymbol{\nu}\neq \mathbf{0}} \frac{1}{|\boldsymbol{\nu}|^{r+1}} = 4K_0 \sqrt{\rho'_{\overline{n}_0-1}} \sum_{\boldsymbol{\nu}\neq\mathbf{0}} \frac{1}{|\boldsymbol{\nu}|^{r+1}}$. Note that a condition like $\tau_1 > \tau + r$ is sufficient to obtain both summability over $\boldsymbol{\nu}$ and a measure (of the excluded set) relatively small with respect to that of I_C. If $\overline{n}_0 < 3$, hence $n_0 < 3$, the same conclusion trivially holds possibly increasing the value of K by a factor 4.

Appendix A3. Resummations: convergence and smoothness

To prove Lemma 2, we first show that the series defining $M^{[n]}(x;\varepsilon)$ for $0 \leq n \leq \overline{n}_0$ converge and then we check smoothness and the bounds. This is done for completeness as the argument is almost a word by word repetition of the analysis in Ref. [GG], with a few slight changes of notations necessary to adapt it to our present notations and scope. To study convergence of the series defining $M^{[n]}(x,\varepsilon)$, $n \leq \overline{n}_0$, we remark that we have to consider only trees in which all propagators have scales $[p]$ with $p \leq \overline{n}_0$. Therefore the propagators which do not vanish will be such that their denominators satisfy $D(x) > 2^{-2(\overline{n}+1)}|x|^2$, see (4.4), so that they are effectively estimated from below by $|x|^2$ times a constant. Note that the case $n = 0$ is obvious (and it is treated in Section 3).

A3.1. *Convergence.* We suppose that the eigenvalues of $\mathcal{M}^{[\leq p]}(x;\varepsilon)$, $n = 0, \ldots, n-1$, differ from the corresponding ones of $\mathcal{M}^{[\leq 0]}(x;\varepsilon) \equiv M_0$ so that $|\lambda_j^{[p]}(x,\varepsilon) - \lambda_j^{[0]}| < \gamma\varepsilon^2$ for some $\gamma > 0$, and that ε is small enough so that $\gamma\varepsilon^2 < \frac{1}{2}\varepsilon a_s 2^{-2\overline{n}-2}$ and, therefore (see (4.4)), if a line with frequency x has scale $[p]$, $p < n$, then $|x^2 - \lambda_j^{[p]}(x,\varepsilon)| > 2^{-2(\overline{n}+2)} x^2$.

We shall use that if a the propagator of a line is on a scale $[n]$ then one has $D(x) \leq 2^{-2(n-2)} C_0^2$, even though *we could allow* also a bound $D(x) \leq 2^{-2(n-1)} C_0^2$. The reason for this is again for later use in bounds necessary to establish the needed cancellations as commented in Section A3.2.

Consider a renormalized self-energy cluster $T \in \mathcal{S}^{\mathcal{R}}_{k,n-1}$, and define $\Lambda_m(T) = \{\ell \in \Lambda(T) : n_\ell = m\}$, for $m \leq n-1$, and $\mathcal{P}(T)$ the set of lines (path) connecting the external lines of T.

If $\boldsymbol{\nu}$ is the momentum flowing in the line entering T then the momentum flowing in a line $\ell \in \Lambda(T)$ of scale $[p]$, $p \leq n-1$, will be $\boldsymbol{\nu}^0_\ell + \sigma_\ell \boldsymbol{\nu}$, $\sigma_\ell = 0, 1$, where $\boldsymbol{\nu}^0_\ell$ is the momentum that would flow on ℓ if $\boldsymbol{\nu} = \mathbf{0}$. The corresponding frequency will be $x'_\ell = x^0_\ell + \sigma_\ell x$, with obvious notations.

First of all we shall prove the bound

$$\sum_{\mathfrak{v}\in V(T)} |\boldsymbol{\nu}_{\mathfrak{v}}| \geq 2^{(n-\overline{n}-5)/\tau_0}. \tag{A3.1}$$

for $T \in \mathcal{S}^{\mathcal{R}}_{k,n-1}$. If there is a line $\ell \in \Lambda_{n-1}(T)$ which does not belong to $\mathcal{P}(T)$ then $x_\ell = x^0_\ell$, so that (A3.1) follows from the Diophantine condition on $\boldsymbol{\omega}$. If all lines in $\Lambda_{n-1}(T)$ belong to $\mathcal{P}(T)$ consider the one among them, say ℓ, which is closest to ℓ^2_T, i.e. the entering line of T. Then call T_1 the connected set of nodes and lines between[7] ℓ and ℓ^2_T. If T_1 is a single node $\mathfrak{v}$ then $\boldsymbol{\nu}_{\mathfrak{v}} \neq \mathbf{0}$,

[7] The lines between two lines ℓ_1 and ℓ_2 with $\ell_2 < \ell_1$ are all the lines which precede ℓ_1 but which do not precede ℓ_2 nor coincide with it.

otherwise $\mathfrak{v}$ would be a trivial node; if T_1 is not a single node then by construction all the lines of T_1 have scales strictly smaller than n, hence $x_\ell \neq x$ otherwise T_1 would be a self-energy cluster. In both cases one has $|x_\ell - x| = |x^0_\ell| > C_0 |\sum_{\mathfrak{v}\in V(T_1)} \boldsymbol{\nu}_\mathfrak{v}|^{-\tau_0}$. On the other hand both $D(x)$ and $D(x_\ell)$ must be $\le (C_0 2^{-(n-2)+1})^2$ hence, by (4.4) $|x|, |x_\ell| \le C_0 2^{-n+\overline{n}+3}$, so that $|x - x_\ell| \le C_0 2^{-n+\overline{n}+4}$, and (A3.1) follows also in such a case.

The next task will be to show that the number $\mathcal{N}_m(T)$ of lines on scale $[m]$, with $m \le n-1$, contained in a cluster T is bounded by $\mathcal{N}_m(T) \le \max\{E_m \sum_{\mathfrak{v}\in V(T)} |\boldsymbol{\nu}_\mathfrak{v}| - 1, 0\}$, with $E_m = E\, 2^{-m/\tau_0}$ for a suitably chosen constant E; as it will emerge from the proof one can take $E = 2\, 2^{(\overline{n}+4)/\tau_0}$.

Before considering clusters we adapt to our context the classical bound (Siegel-Bryuno-Pöshel; see Ref. [Ga] and references quoted therein), stating that, if $\mathcal{N}_m(\theta)$ denotes the number of lines on scales $[m]$, then by induction on the number of nodes of θ one shows: $\mathcal{N}_m(\theta) \le \max\{E_m \sum_{\mathfrak{v}\in V(\theta)} |\boldsymbol{\nu}_\mathfrak{v}| - 1, 0\}$. Indeed if θ contains only one node $\mathfrak{v}_0$ and the frequency $x = \boldsymbol{\omega}\cdot\boldsymbol{\nu}_{\mathfrak{v}_0}$ of the root line has scale $[m]$ one has

$$2^{-m+1} C_0 \ge \sqrt{D(x)} \ge 2^{-(\overline{n}+1)} |x| \ge 2^{-(\overline{n}+1)} C_0 |\boldsymbol{\nu}_{\mathfrak{v}_0}|^{-\tau_0} \Rightarrow |\boldsymbol{\nu}_{\mathfrak{v}_0}| > 2^{(m-\overline{n}-2)/\tau_0}, \qquad (A3.2)$$

hence $E_m |\boldsymbol{\nu}_{\mathfrak{v}_0}| - 1 \ge 2$ and the bound holds in this simple case.

If θ has k nodes and the root line *does not* have scale $[m]$ the inductive assumption, if it is assumed for the cases of $k' < k$ nodes, gives the bound for k-nodes trees.

If the root line has scale $[m]$ then on each path of tree lines leading to the root we select the line among the ones on scales $[m']$ with $m' \ge m$ *closest to the root* (if any is found on the path) and we call the selected lines $\ell_1, \ldots, \ell_q$. If $q \neq 1$ either the bound follows just as in the case of $k = 1$ (when $q = 0$) or from the inductive hypothesis (when $q \ge 2$).

The case $q = 1$ and $[n_{\ell_1}] = [\infty]$ (i.e. $\boldsymbol{\nu}_{\ell_1} = \mathbf{0}$) can be treated as the case $q = 0$. If $q = 1$ and $\boldsymbol{\nu}_{\ell_1} \neq \mathbf{0}$, by construction all lines between the root line ℓ and ℓ_1, see footnote [7], have scales $[m']$, with $m' < m$, so that such lines, together with the nodes they connect, form a cluster T. The frequencies x_ℓ and x_{ℓ_1} must be different because the tree θ contains no self-energy clusters. On the other hand $\sqrt{D(x_\ell)}, \sqrt{D(x_{\ell_1})} \le 2^{-m+1} C_0$, hence $|x_\ell|, |x_{\ell_1}| \le 2^{-m+\overline{n}+3} C_0$ by (4.4), and $C_0 |\boldsymbol{\nu}_\ell - \boldsymbol{\nu}_{\ell_1}|^{-\tau_0} \le |x_\ell - x_{\ell_1}| \le 2^{-m+\overline{n}+4} C_0$, so that we get $\sum_{\mathfrak{v}\in V(T)} |\boldsymbol{\nu}_\mathfrak{v}| \ge (2^{-m+\overline{n}+4})^{-1/\tau_0}$, which gives $\mathcal{N}_m(\theta) \le 1 + E_m \sum_{\mathfrak{v}\in V(\theta)} |\boldsymbol{\nu}_\mathfrak{v}| - E_m \sum_{\mathfrak{v}\in V(T)} |\boldsymbol{\nu}_\mathfrak{v}| - 1 \le E_m \sum_{\mathfrak{v}\in V(\theta)} |\boldsymbol{\nu}_\mathfrak{v}| - 1$, so that the bounds is completely proved.

Remark. The above discussion exploits the property that the tree θ that we consider *cannot*, by definition of renormalized tree, contain self-energy clusters, and follows Ref. [GG] which was based on the possibility of bounding the denominators proportionally to x^2 (in that case the proportionality factor was 1): a property also valid here for $n \le \overline{n}_0$.

For the bound on $\mathcal{N}_m(T)$ we consider a subset G_0 of the lines of a tree θ between two lines $\ell_{\rm out}$ and $\ell_{\rm in}$. Set $G = G_0 \cup \ell_{\rm in} \cup \ell_{\rm out}$. Let $[p_{\rm in}], [p_{\rm out}]$ be the scales of the lines $\ell_{\rm out}$ and $\ell_{\rm in}$, respectively, and suppose that $p_{\rm in}, p_{\rm out} \ge m$, while all lines in G_0 (if any) have scales $[p]$ with $p \le n-1$. Note that in general G_0 is not even a cluster unless $p_{\rm in}, p_{\rm out} \ge n$. Then we can prove that $\mathcal{N}_m(G_0) \le \max\{E_m \sum_{\mathfrak{v}\in V(G_0)} |\boldsymbol{\nu}_\mathfrak{v}| - 1, 0\}$, where $V(G_0)$ is the set of nodes preceding $\ell_{\rm out}$ and following $\ell_{\rm in}$, and E_m is defined above. If G_0 has zero lines then the harmonic $\boldsymbol{\nu}_0$ of the (only) node in $V(G_0)$ is large, $|\boldsymbol{\nu}_0| \ge 2^{(m-\overline{n}-2)/\tau_0}$ (by the Diophantine property) and the statement is true. Hence we proceed inductively on the number of lines in G_0.

If no line of G_0 on the path $\mathcal{P}(G)$ connecting the external lines of G has scale $[m]$ then the lines in G_0 on scale $[m]$ (if any) belong to trees with root on $\mathcal{P}(G)$, and the statement follows from the bound on trees.

Suppose that $\ell\in\mathcal{P}(G)$ is a line on scale $[m]$, then call G_1 and G_2 the disjoint subsets of G such that $G_1\cup G_2\cup\ell=G$. Then $G_1\cup\ell$ and $G_2\cup\ell$ have the same structure of G itself but each has less lines: and again the inductive assumption yields the result.

Therefore, as a particular case, by choosing $G_0=T$, with $T\in\mathcal{S}^{\mathcal{R}}_{k,n-1}$, the bound for $\mathcal{N}_m(G)$ implies the bound on $\mathcal{N}_m(T)$ we are looking for.

The above analysis is taken from Ref. [Ge] and differs from Ref. [GG] because here the scales depend on ε and it is not clear how to define a "strong Diophantine condition", which would allow a one-to-one correspondence between line scales and line momenta.

The bound on the contribution of a single self-energy cluster $T\in\mathcal{S}^{\mathcal{R}}_{k,n-1}$ is then

$$\begin{aligned}&\frac{\varepsilon^k}{k!}C_0^{-2k}F^{2k}e^{-\frac12\kappa_0\sum_{\mathfrak{v}}|\boldsymbol{\nu}_{\mathfrak{v}}|}\Big(\prod_{m=0}^{m_0}2^{(m+3)2\mathcal{N}_m(T)}\Big)\cdot\\ &\cdot\Big(e^{-\frac12\kappa_0\sum_{\mathfrak{v}\in V(T)}|\boldsymbol{\nu}_{\mathfrak{v}}|}\prod_{m=m_0+1}^{\infty}2^{2(m+3)2^{-m/\tau_0}E\sum_{\mathfrak{v}\in V(T)}|\boldsymbol{\nu}_{\mathfrak{v}}|}\Big)\;\le\;\frac{\varepsilon^\kappa G^k}{k!}e^{-\frac12\kappa_0\sum_{\mathfrak{v}\in V(T)}|\boldsymbol{\nu}_{\mathfrak{v}}|},\end{aligned}\tag{A3.3}$$

with F an upper bound on the constants F_0,F_1 bounding the Fourier transform of the perturbation (see (1.4)), while m_0 is defined so that $\log 2\sum_{m>m_0}2(m+3)2^{-m/\tau_0}E\le\frac12\kappa_0$ and G is a suitable constant.

The number of trees can be bounded by $4^kk!$, and the sum over the scale labels involves at most 2 possible values per line because of the upper and lower cut-offs present in the propagators definition. The sum over the harmonics can be estimated by making use of *part* of the exponential factor in (A3.3) (say $\frac14\kappa_0$) while the other $\frac14\kappa_0$ will be used as a factor bounded by $e^{-\frac14\kappa_0 2^{(n-\overline{n}-5)/\tau_0}}$, by (A3.1).

Hence we get convergence at exponential rate 2^{-1} for $\varepsilon<\varepsilon_1$ (and ε_1 is an explicitly computable constant) and the matrix $M^{[n]}(x;\varepsilon)$ is defined by a convergent series and it is bounded by

$$\|M^{[n]}(x;\varepsilon)\|<\overline{B}\varepsilon^2e^{-\frac14\kappa_0 2^{(n-\overline{n}-4)/\tau_0}},\tag{A3.4}$$

for a suitable $\overline{B}$ which can be read from (A3.3), i.e. we get the first of the first line in (5.13) with the constant B replaced by $\overline{B}$, $\tau=\tau_0$, and $\kappa_1=\frac14\kappa_0e^{-(\overline{n}+4)/\tau_0}$. The ε^2 factor is due to the parallel remark that, in any self-energy cluster whose value contributes to $\mathcal{M}^{[n]}(x;\varepsilon)$, k is certainly ≥ 2 (see Remark to Definition 4 in Section 5).

Therefore if ε is small enough (that is smaller than a constant independent of $n\le\overline{n}_0$)

$$\|\mathcal{M}^{[\le n]}(x;\varepsilon)-M_0\|\le\overline{B}\varepsilon^2\sum_{n=1}^{\infty}e^{-\frac14\kappa_0 2^{(n-\overline{n}-4)/\tau_0}}\stackrel{def}{=}B'\varepsilon^2,\tag{A3.5}$$

so that the eigenvalues of $\mathcal{M}^{[\le n]}(x;\varepsilon)$ will be shifted with respect to the corresponding eigenvalues of M_0 by $\gamma\varepsilon^2$ at most, with $\gamma\stackrel{def}{=}B'C$, see (I) in Appendix A4.

Hence if we define γ as $B'C$ and ε is chosen small enough, say $\varepsilon<\varepsilon_2$, so that $\gamma\varepsilon^2<\frac12\varepsilon a_s2^{-2\overline{n}-2}$ (as it must be in order that the above argument be consistent, see the beginning of the current

Section) we obtain the validity of the assumed inductive hypothesis for all $n \le \overline{n}_0$ and of the first inequality in the first line of (5.13) where B can be chosen equal to $\overline{B}$ above.

The symmetries in items (i) and (ii) are an algebraic consequence of the form of the Lindstedt series: hence they are a necessary consequence of the proved convergence, see Ref. [GG].

A3.2. *Smoothness.* The function $M^{[n]}(x;\varepsilon)$ which we have just shown to be well defined for all ε small enough will be smooth in ε, x. We assume inductively that this is the case for $M^{[p]}(x;\varepsilon)$, $0 \le p < n-1$, and that the bounds in the first line of (5.13) hold for such p's (the case $p=0$ is obvious as $\mathcal{M}^{[0]}(x;\varepsilon) \equiv M_0$).

Each derivative with respect to x or, respectively, to ε will replace the value of a self-energy cluster with k nodes by a sum of k terms which can be bounded by a bound like (A3.3).

In fact, given a self-energy cluster T, the right derivative ∂_x^+ may fall on a denominator of one of the $k-1$ cluster lines. If its frequency is $x+x_0$ with scale label $[m]$, derivation yields, up to a sign, a product of two matrices $((x_0+x)^2 - \mathcal{M}^{[\le m]}(x_0+x;\varepsilon))^{-1}$ times $2\,(x_0+x) - \partial_x^{\pm}\mathcal{M}^{[\le m]}(x_0+x;\varepsilon)$ with an appropriate order of multiplication. The term $2\,(x+x_0)\,((x_0+x)^2 - \mathcal{M}^{[\le m]}(x_0+x;\varepsilon))^{-2}$ can be bounded proportionally to $(C_0^{-2}2^{2(m-1)})^{3/2} \le (C_0^{-2}2^{2(m-1)})^2$, while the remaining term can be studied by making use of the inductive assumption $\|\partial_x\mathcal{M}^{[\le m]}(x_0+x;\varepsilon)\| \le B\varepsilon^2 a_s^{-1/2}$ and it leads to the same bound found for the first term, i.e. $(C_0^{-2}2^{2(m-1)})^2$, *multiplied* by $B\varepsilon^2$.[8]

If the derivative falls on either a ψ_p or a χ_p function, we can use that such derivative can be bounded proportionally to $C_0^{-1}2^p$ and $\sum_{p=0}^{m-1} 2^p = 2^m$, to obtain again the same bound as the first case.

Hence the final bound has the form $B_1 + \varepsilon^2 Bb$ with B_1, b suitable constants, provided ε is small enough, say $\varepsilon < \varepsilon_3$. The value of the constants B_1, b do not depend on the inductively assumed value for B: in particular B_1 can be obtained (see Remark (2) below for a smarter bound) by replacing $2^{(m+3)}$ in the two factors in the *l.h.s.* of (A3.3) by $2^{2(m+3)}$ and by inserting a factor k times a constant (to keep track of all the constant factors arising from differentiation). Therefore if $B = 2B_1$ the estimate on $\partial_x^+\mathcal{M}^{[\le n]}(x;\varepsilon)$ follows if ε is small enough, say $\varepsilon < \varepsilon_4$. The same can be said about the left derivative ∂_x^-.

The right and left differentiability of $\mathcal{M}^{[n]}(x;\varepsilon)$ with respect to x is due to the dependence of $\mathcal{M}^{[n]}(x;\varepsilon)$ on the function $D(x)$: the latter has a discontinuous derivative at a finite number of points (roughly at midpoints between the eigenvalues $\lambda_j^{[0]}$ of M_0).[9] Note that the denominators in the self-energy values defining $M^{[n]}(x;\varepsilon)$ cannot vanish, and actually stay well away from 0, permitting the above bounds, because of the lower cut-off $\psi_0(D(x))$ appearing in the definition of the propagators $g^{[0]}(x;\varepsilon)(x;\varepsilon)$; see (5.6) and (5.7).

The same argument holds for ∂_ε: however the bound will be only $B\varepsilon$ instead of $B\varepsilon^2$ because the derivative with respect to ε might decrease by one unit the degree of the self-energy values involved. Thus the first line of (5.13) is completely proved. Of course for each of the three terms we get a different constant B, but for simplicity we use for them all the largest, still calling it B.

Remark. (1) We could also prove existence of higher x,ε-derivatives of $\mathcal{M}^{[n]}(x;\varepsilon)$ and of its

[8] Since the matrix $M^{[m]}(x_0+x;\varepsilon)$ is generated by self-energy clusters clusters of degree at least 2.

[9] One could avoid having only left and right differentiability by using a regularized version of the function $D(x)$ as discussed in Remark (2) after Lemma 2 in Section 5.

eigenvalues $\lambda_j^{[n]}(x,\varepsilon)$ for $j>r$ via the above argument.
(2) *The more derivatives we try to estimate with the above method the smaller would become the set of allowed values of* ε. This however is avoidable. Instead of imagining to include the bound $C_0^{-2}2^{2m}$ arising above as a consequence of the "extra" $D(x+x_0)$ or of the other derivatives into the factors 2^{m+3} associated with the divisors in (A3.3) one could simply further bound this by $C_0^{-2}2^{2n}$ and use part of the factor $e^{-\frac14\kappa_0 2^{(n-\overline{n}-4)/\tau_0}}$ (say replacing $\frac14\kappa_0$ with $\frac18\kappa_0$): this eventually leads to a bound on the s-th right-derivative with respect to x of a value of a self-energy cluster proportional to $2^{ns}e^{-\frac18\kappa_0 2^{(n-\overline{n}-4)/\tau_0}}$ but with an s-independent estimate of the radius of convergence (as the constant G in (A3.3) remains the same). This is sufficient to get the existence of the s-th derivatives without any further restriction on ε: and a similar argument holds for the ε-derivatives.

A3.3. *Cancellations.* Only the bound in the fourth line of (5.13) follows from those in the first line. The bounds in the second and third lines express remarkable properties of Lindstedt series and are essentially algebraic properties: they are the "same" cancellations which occur in KAM theory, see Refs. [Ga], [GM1], and are based on the remark that if T is a self-energy cluster the entering and exiting lines have the same momentum $\boldsymbol{\nu}$: hence the sum of the harmonics of the nodes of T vanishes $\sum_{\mathfrak{v}\in V(T)}\boldsymbol{\nu}_{\mathfrak{v}}=\mathbf{0}$.

We start by dealing with the trivial cases. Consider first self-energy clusters T such that

$$\sum_{\mathfrak{v}\in V(T)}|\boldsymbol{\nu}_{\mathfrak{v}}|\ge (C_0/2^6|x|)^{1/\tau_0}. \tag{A3.6}$$

For such a self-energy cluster T one can use part (say 1/8) of the exponential decay of the node factors to obtain a bound $e^{-\frac{\kappa_0}{8}\sum_{\mathfrak{v}\in V(T)}|\boldsymbol{\nu}_{\mathfrak{v}}|}\le e^{-b_1|x|^{-1/2\tau_1}}\le b_2x^2$, with b_1 and b_2 two suitable positive constants, while a factor ε^2 simply follows from the fact that any self-energy cluster has at least two nodes.

So we can assume that (A3.6) *does not hold.* If $\boldsymbol{\nu}$ is the momentum flowing in the entering line then the momentum flowing in a line $\ell\in\Lambda(T)$ of scale $[p]$, $p\le n$, if the scale of the cluster is $[n]$, will be $\boldsymbol{\nu}_\ell^0+\sigma_\ell\boldsymbol{\nu}$, $\sigma_\ell=0,1$, where $\boldsymbol{\nu}_\ell^0$ is the momentum that would flow on ℓ if $\boldsymbol{\nu}=\mathbf{0}$. The corresponding frequency will be $x_\ell=x_\ell^0+\sigma_\ell x$, with obvious notations.

If the entering and exiting lines are imagined attached to the internal nodes of T in all possible ways (i.e. in k^2 ways if T contains k nodes) *keeping all their labels unaltered* then one obtains a family $\mathcal{F}_T$ of self-energy clusters. The contribution of each self-energy cluster of $\mathcal{F}_T$ to each of the entries of the matrix $\mathcal{M}^{[n]}(x;\varepsilon)$ with labels $i,j\le r$ (the $\alpha\alpha$ entries in the notations of Lemma 2) and with labels $i\le r, j>r$ (the $\alpha\beta$ entries) has the form $M_{i,j;\mathfrak{v},\mathfrak{w}}(x,T)\,\nu_{\mathfrak{v},i}\,\nu_{\mathfrak{w},j}$ or, respectively, $M'_{i,j;\mathfrak{v}}(x,T)\,\nu_{\mathfrak{v},i}$, with

$$\begin{aligned} &M_{i,j;\mathfrak{v},\mathfrak{w}}(x,T)=M_{i,j}(T)+xM^{(1)}_{i,j}(T)+x^2M^{(2)}_{i,j,\mathfrak{v},\mathfrak{w}}(x,T), && i,j\le r,\\ &M'_{i,j;\mathfrak{v}}(x,T)=M'_{i,j}(T)+xM'^{(1)}_{i,j,\mathfrak{v}}(x,T), && i\le r,\ r<j, \end{aligned} \tag{A3.7}$$

so that after performing the sum over the self-energy clusters of $\mathcal{F}_T$, i.e. after performing the sums $\sum_{\mathfrak{v},\mathfrak{w}\in V(T)}$ or, respectively, $\sum_{\mathfrak{v}\in V(T)}$, the first two terms in the first line and the first term in the second line do not contribute because $\sum_{\mathfrak{v}}\boldsymbol{\nu}_{\mathfrak{v}}=\mathbf{0}$. However one has to show that the matrices M and M' in the *r.h.s.* of (A3.7) satisfy appropriate bounds once the factors x determining the order of zero at $x=0$ are extracted. From the convergence one expects that the bounds should still be

proportional to ε^2 while the derivatives $\partial_x^\pm$ or ∂_ε should satisfy bounds proportional to ε^2 or to ε respectively.

The (A3.7) are proved by means of interpolations, see [GM1], between the contributions of the self-energy clusters in the family $\mathcal{F}_T$. When we collect together the values of the self-energy clusters in $\mathcal{F}_T$ then the arguments of some of the propagators can fall outside the supports of the respective cut-off function (because the lines are shifted but their scale labels are kept fixed so that scales of the propagators of the self-energy clusters $T' \in \mathcal{F}_T$ are the ones inherited by T while the momentum flowing in them may change).

This generates trees and clusters for which we made no estimates (because they are just 0). However when interpolating we may end up computing values of trees, with scale assignments which would give a value 0, at *intermediate* frequencies where the values no longer vanish. In estimating such interpolated values we can proceed as in the cases already treated, but it will not be necessarily true that a line of frequency x and scale $[n]$ will satisfy $2^{-2(n-1)}C_0^2 < D(x) \leq 2^{-2n}C_0^2$. Nevertheless a slightly weaker version of this inequality has to hold in which the the *l.h.s.* is divided by 4 and the *r.h.s.* is multiplied by 4 (cf. also Remark (3) after the inductive assumption in Section 6), and the estimates will not only be possible but they can be regarded as already obtained because, as the reader can check, we have been careful in discussing the bounds obtained so far under such weaker condition. This also clarifies why we have defined $\overline{n}$ in (4.2) one unit larger than what appeared there as necessary so that the estimate (4.4) is apparently worse than it should.

In some cases, however, a serious problem seems to arise when actually attempting to derive bounds: namely the bounds on the matrices which appear as coefficients in (A3.7) can really be checked as just outlined by the above hints, and without affecting the values of ε for which one has convergence, only if x verifies the condition of being so small that the variations of the momenta flowing in the inner lines of T, when the entering or exiting lines are moved and re-attached to all nodes of T, remain so small that the quantities $D(x_\ell)$ corresponding to the lines ℓ in the cluster T stay essentially unchanged.

In certain cases shifting the entering or exiting lines to the nodes of the self-energy cluster T may considerably change the scales of the lines ℓ in T, but this is the case in which (A3.6) holds. And precisely in such a case the cancellations are not needed to prove the bound, because we have checked that the value of *each* self-energy cluster contributing to $M^{[n]}$ individually already verifies that bound that we want to prove.

If (A3.6) does not hold, then two cases are possible: either $|x|$ is close to $\lambda_j^{[p]}$ for some $j > r$ or larger, and no cancellation occurs, or $|x|$ is $< C_0 2^{-n}$. In the latter case the inequality opposite to (A3.6) implies that for $\ell \in \Lambda(T)$ one has $|x_\ell^0| \geq 4|x|$, hence $2|x_\ell^0| \geq |x_\ell| \geq \frac{1}{2}|x_\ell^0|$, so that the scales can change by at most one unit by shifting the external lines of T. Then the quantities $D(x_\ell)$ do not change much for all lines $\ell \in \Lambda(T)$, and we shall have the cancellation through the mentioned mechanism. Therefore the contribution of $\mathcal{M}^{[p]}(x;\varepsilon)$ to $\mathcal{M}^{[\leq n]}(x;\varepsilon)$ can be bounded in both cases proportionally to $e^{-\frac{1}{4}\kappa_0 2^{(n-\overline{n}-4)/\tau_0}}$ times $\min\{\varepsilon^2, \varepsilon\,|x|^2\}$ for the entries $\alpha\alpha$ or times $\min\{\varepsilon^2, \varepsilon^{\frac{3}{2}}|x|\}$ for the $\alpha\beta$ entries: either by the cancellation (second case) or by the general bound $O(\varepsilon^2)$ on matrix elements (first case), because x^2 is of order $O(\varepsilon)$.

Finally we note that in the estimates of the M's in (A3.7) we have to sum over the scale labels and this gives a factor per line larger than the one arising in the bound (A3.3) (which was 2); in fact we have to consider also trees with vanishing value: but the scales of the divisors associated with their lines can change at most by one unit with respect to the scale, hence we can have at

most 4 scale labels per line.

Remark. We stress once more that the above analysis holds if ε is small enough, say $\varepsilon < \overline{\varepsilon}_1$ with $\overline{\varepsilon}_1$ determined by collecting all the (three) restrictions imposed by requiring ε to be "small enough", derived above and $\overline{\varepsilon}_1$ is *independent* of n_0 (otherwise it would be uninteresting). The reason is that as long as we do not deal with x's which are too close to the eigenvalues of M_0, so that the key inequality (4.4) holds, we do not really see the difference between the hyperbolic and the elliptic cases: and in the hyperbolic cases there is no need for a lower cut-off at scale $\sim n_0$ where resonances between the proper frequencies (which are of order ε) and the elliptic normal frequencies become possible (as $\varepsilon \simeq C_0^2 2^{-2n_0}$).

A3.4. *Resonant resummations.* Concerning the proof of Lemma 6 we only need to add a few comments. The bounds on $\mathcal{N}_m(\theta)$ and $\mathcal{N}_m(T)$ can be discussed exactly as for the scales $[n]$ with $n \le \overline{n}_0$, with the only difference that now one has to use also the second part of the Diophantine conditions (6.7), as already done in the argument leading to (6.12); in particular the role of the exponent τ_0 is now played by $2\tau_1$ (because of the Diophantine conditions in (6.7) which replaces (1.3) in the discussion), while in the analogues of (A3.1) and the following bounds no $\overline{n}$ appear, as the propagator divisors are bounded directly in terms of the corresponding scales, and not in terms of the frequencies.

Also the argument given above about the cancellations extends easily to the scales $[n]$, with $n \ge \overline{n}_0$. The only difference is that in (A3.6) the exponent $1/\tau_0$ has to be replaced with $1/(2\tau_1)$, in such a way that for any line $\ell \in \Lambda(T)$ one has $||x_\ell^0| - \sqrt{|\lambda_j^{[n_\ell-1]}(\varepsilon)|}| \ge 4|x|$, hence the chain of inequalities

$$2\left||x_\ell^0| - \sqrt{|\lambda_j^{[n_\ell-1]}(\varepsilon)|}\right| \ge \left||x_\ell| - \sqrt{|\lambda_j^{[n_\ell-1]}(\varepsilon)|}\right| \ge \frac{1}{2}\left||x_\ell^0| - \sqrt{|\lambda_j^{[n_\ell-1]}(\varepsilon)|}\right|, \tag{A3.8}$$

follows, and again by shifting the external lines of T the scales of the internal lines can change at most by one unit, when (A3.6) is not satisfied

Appendix A4. Matrix properties

(I) *Let M_0 be a $d\times d$ Hermitian matrix with eigenvalues $\lambda_1 < \ldots < \lambda_p$ with multiplicities $n_1, \ldots, n_p$ and eigenspaces $\Pi_1, \ldots, \Pi_p$ on which we fix orthonormal bases $\underline{e}_{j,k}$, $j = 1, \ldots, p, k = 1, \ldots, n_j$. Let M_1 be Hermitian and consider the matrix $M = M_0 + \varepsilon M_1$. There exists a constant C such that, for ε small enough, there will be n_j eigenvalues of M (not necessarily all distinct) which are analytic in ε and one has $|\lambda_{j,k}(\varepsilon) - \lambda_{j,k'}(\varepsilon)| \le C\varepsilon$ for $k, k' = 1, \ldots, n_j$.*

Hint. If $n_j = 1$ this follows immediately form the formula $\lambda_j(\varepsilon) = \mathrm{Tr}\left(\frac{1}{2\pi i}\oint_{\gamma_j}\frac{z dz}{z-M}\right)$, where γ_j is a circle around $\lambda_j(0)$ of ε-independent radius smaller than half the minimum separation δ between the λ_j for ε small enough (so that $C_1\varepsilon^{\frac{1}{d}} < \delta$ for a suitable C_1)[10].

[10] Because the characteristic polynomials $P(\lambda), P_0(\lambda)$ are related by $P(\lambda) = P_0(\lambda) + \varepsilon Q(\lambda, \varepsilon)$ with Q of lower degree. Therefore there is L such that if $|\lambda| > L$ then for all $|\varepsilon| < 1$ (say) it is $P(\lambda) \ne 0$. Furthermore if all roots of P differ by at least y from those of P_0 one has $|P(\lambda)| \ge y^d - \varepsilon C^d$ where $C^d = \max_{|\Lambda|\le L, |\varepsilon|\le 1}|Q(\lambda, \varepsilon)$. Hence $y \le C\varepsilon^{d^{-1}}$.

Otherwise it follows form similar formulae for the projection operator E_j on Π_j and for $E_j M E_j$:

$$E_j = \frac{1}{2\pi i}\oint_{\gamma_j} \frac{dz}{z-M}, \qquad E_j M E_j = \frac{1}{2\pi i}\oint_{\gamma_j} E_j \frac{z\,dz}{z-M} E_j, \tag{A4.1}$$

which, for ε small, can be expanded into a convergent power series in ε (as done explicitly in a similar context in (A4.3) below) because of the ε-independence of the radii of γ_j. One can also construct an orthonormal basis on Π_j with vectors of the form $\mathbf{v}_{j,k} = \mathbf{e}_{j,k} + \sum_{q\ge 1}^{\infty} \varepsilon^q \mathbf{e}^{(q)}_{j,k}$ (simply applying the Hilbert-Schmidt orthonormalization to the vectors $E_j \mathbf{e}_{j,k}, k = 1, \ldots, n_j$). One then remarks that the matrix $E_j M E_j$ has n_j eigenvalues and that it has the form $\lambda_j + \varepsilon \widetilde{M}(\varepsilon)$.

So the problem is reduced to the case in which M_0 is the identity perturbed by an analytic matrix. Either $\widetilde{M}(\varepsilon)$ is proportional to the identity and there is nothing more to do, or it is not: hence there will be an order in ε at which the degeneracy is removed and repeating the argument we reduce the problem to a similar one for matrices of dimension lower than n_j: and so on until we find a matrix (possibly one dimensional) proportional to the identity to all orders. ∎

In our analysis we need the following corollary.

(II) *Let M_0 be Hermitian with r degenerate eigenvalues equal to 0 and $s = d-r$ simple eigenvalues εa_j, $j = 1, \ldots, s$.*
(i) The matrix $M_0 + \varepsilon^2 M_1$ with M_1 Hermitian and differentiable in ε with bounded derivative will have s non-degenerate eigenvalues $\varepsilon a_j + O(\varepsilon^2)$, $j = 1, \ldots, s$, and r eigenvalues $\lambda_1(\varepsilon), \ldots, \lambda_r(\varepsilon)$, all analytic in ε, with the property that for all $k = 1, \ldots, r$ one has $|\lambda_k(\varepsilon)| < C\,\varepsilon^2$, if ε is small enough and C is a suitable constant.
(ii) If M_1 depends on a parameter x and is differentiable also in x with bounded derivative then

$$\begin{aligned} &|\partial_x \lambda_j(x;\varepsilon)| \le C\varepsilon^2, \quad |\partial_\varepsilon \lambda_j(x;\varepsilon)| \le C, && j > r, \\ &|\lambda_j(x;\varepsilon) - \lambda_j(x';\varepsilon)| \le C\varepsilon^2 |x-x'|^{1/r}, && j \le r, \end{aligned} \tag{A4.2}$$

if ε is small enough and C is a suitable constant.

The second relation in (A4.2) is not used in this paper and is given only for completeness.

Hint. We apply the previous lemma to the matrices $\varepsilon^{-1} M_0$ and εM_1 and we get (i). To get (A4.2) we note that the x-derivative of $M_0 + \varepsilon^2 M_1$ is $\varepsilon^2 \partial_x M_1$ and the first of (A4.2) follows. To obtain the second of (A4.2) we have to compare the eigenvalues of $M_0 + \varepsilon^2 M_1(x;\varepsilon)$ with those of $M_0 + \varepsilon^2 M_1(x;\varepsilon) + \varepsilon^2 O(|x-x'|)$. By the above expression for the projection on the plane of the first r eigenvalues this is reduced to the problem of comparing two $r \times r$ matrices of order ε^2 and differing by $O(|x-x'|)$. The power $1/r$ arises from the estimate that the considered projection of the matrix M_1 (which is only differentiable in x) has r eigenvalues close to 0 within $C_1 |x-x'|^{1/r}$, for some $C_1 > 0$ (by (I) above), and ε is small enough. Hence we get the second of (A4.2). ∎

A third property that we need is the following one.

(III) *If M_0 is as in (II) and M_1 is Hermitian and has the form $\begin{pmatrix} \varepsilon^2 x^2 N & \varepsilon^2 x P \\ \varepsilon^2 x P^* & \varepsilon^2 Q \end{pmatrix}$, with N, Q two $r \times r$ and $r \times s$ matrices and P a $r \times s$ matrix then the first r eigenvalues of $M_0 + M_1$ are bounded*

by $|\lambda_j(x,\varepsilon)| < C\varepsilon^2 x^2$*, for* $j = 1,\ldots,r$.

Hint. This is obtained by using (A4.1) which gives the projection over the plane of the r eigenvalues within $O(\varepsilon^2)$ of 0 as integral over a circle of radius $\frac{1}{2}a_1\varepsilon$

$$E = \frac{1}{2\pi i}\oint_\gamma \frac{dz}{z-M_0}\sum_{k=0}^{\infty}(M_1\frac{1}{z-M_0})^k, \tag{A4.3}$$

and one sees that $(M_1\frac{1}{z-M_0})^k$ has for all $k\ge 1$ the same form of M_1, with ε^2 replaced by ε^{2k}, so that the sum of the series is the matrix $\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$ corresponding to the $k=0$ term (it is a $d\times d$ block matrix with the first $r\times r$ block 1 and the other blocks 0) plus a matrix of the same form of M_1. Likewise the basis $\mathbf{v}_h = E\mathbf{e}_h$, $h=1,\ldots,r$ consists of vectors of the form $\mathbf{e}_h + \begin{pmatrix}\varepsilon^2x^2\mathbf{u}_h\\ \varepsilon^2 x\mathbf{u}'_h\end{pmatrix}$, so that one checks that the matrix $(\mathbf{v}_h,(M_0+M_1)\mathbf{v}_{h'})$ is a $r\times r$ matrix which is proportional to ε^2x^2 (i.e. it has the form $\varepsilon^2x^2M_2(x,\varepsilon)$, with M_2 bounded for ε small and for $|x|<1$) and which, by construction, has the same eigenvalues as the first r eigenvalues of the matrix M_0+M_1.

For the above properties see also [RS] and [Ka].

Appendix A5. Algebraic identities for the renormalized expansion

We show that the function $\mathbf{h}$ defined through the renormalized expansion solves the equations of motion (1.5) for all $\varepsilon\in\mathcal{E}$. This is essentially a repetition of Ref. [Ge]. We shall check that $\mathbf{h} = \varepsilon g\partial_{\boldsymbol{\varphi}} f(\boldsymbol{\psi}+\mathbf{a},\boldsymbol{\beta}_0+\mathbf{b})$, where $\boldsymbol{\varphi} = (\boldsymbol{\alpha},\boldsymbol{\beta})$ and g is the pseudo-differential operator with kernel $g(\boldsymbol{\omega}\cdot\boldsymbol{\nu}) = (\boldsymbol{\omega}\cdot\boldsymbol{\nu})^{-2}$. We can write $\mathbf{h} = \sum_{\boldsymbol{\nu}\in\mathbb{Z}^r} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}}\mathbf{h}_{\boldsymbol{\nu}}$, $\mathbf{h}_{\boldsymbol{\nu}} = \sum_{n=0}^{\infty}\mathbf{h}_{n,\boldsymbol{\nu}}$ (only two terms in this series are different from 0 for each $\boldsymbol{\nu}$), with $\mathbf{h}_{n,\boldsymbol{\nu}} = \sum_{k=1}^{\infty}\sum_{\theta\in\Theta^{\mathcal{R}}_{k,\nu}(n)}\mathrm{Val}(\theta)$, where $\Theta^{\mathcal{R}}_{k,\boldsymbol{\nu}}(n)$ is the set of trees in $\Theta^{\mathcal{R}}_{k,\boldsymbol{\nu}}$ such that the root line has scale n. With respect to the previous sections we have dropped the component label $\gamma\in\{1,\ldots,d\}$ in the definition of the set of trees, for notational convenience.

Note that, for all $x\neq 0$ and for all $p\ge 0$ one has

$$1 = \sum_{n=p}^{\infty}\psi_n(\Delta^{[n]}(x,\varepsilon))\prod_{q=p}^{n-1}\chi_q(\Delta^{[q]}(x,\varepsilon)), \tag{A5.1}$$

where the term with $n=p$ has to be interpreted as $\psi_p(\Delta^{[p]}(x;\varepsilon))$.

Set $\Psi_n(x;\varepsilon) = \psi_n(\Delta^{[n]}(x;\varepsilon))\prod_{p=0}^{n-1}\chi_p(\Delta^{[p]}(x;\varepsilon))$ for $n\ge 1$, $\Psi_0(x;\varepsilon) = \psi_0(\Delta^{[0]}(x;\varepsilon))$: by using

(A5.1) one can write, in Fourier space and evaluating the functions of $\boldsymbol{\varphi}$ at $\boldsymbol{\varphi}=(\boldsymbol{\psi}+\boldsymbol{\alpha},\boldsymbol{\beta}_0+\boldsymbol{\beta})$,

$$
\begin{aligned}
g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} &= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} \\
&= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)(g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))^{-1}g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} \\
&= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} \\
&= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)\sum_{k=1}^{\infty}\sum_{\theta\in\overline{\Theta}^{\mathcal{R}}_{k,\nu}(n)}\mathrm{Val}(\theta),
\end{aligned}
\tag{A5.2}
$$

where $\overline{\Theta}^{\mathcal{R}}_{k,\boldsymbol{\nu}}(n)$ differs from $\Theta^{\mathcal{R}}_{k,\boldsymbol{\nu}}(n)$ *as it contains also trees which can have one renormalized self-energy cluster* T *with exiting line* ℓ_0, if ℓ_0 denotes the root line of θ; for such trees the line entering T will be on a scale $p\ge 0$, while the renormalized self-energy cluster T will have a scale $n_T=q$, with $q+1\le\min\{n,p\}$.

Remark. Note that in both (A5.1) and (A5.2) only a finite number of addends is different from zero, as the analysis of Section 6 shows, so that the two series are well defined. The same observation applies to the following formulae, where appear series which, in fact, are finite sums.

By explicitly separating in (A5.2) the trees containing such self-energy clusters from the others,

$$
\begin{aligned}
g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} &= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)\sum_{k=1}^{\infty}\sum_{\theta\in\Theta^{\mathcal{R}}_{k,\nu}(n)}\mathrm{Val}(\theta) \\
&+ g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=1}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon) \\
&\qquad \sum_{p=n}^{\infty}\sum_{q=0}^{n-1}M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\sum_{k=1}^{\infty}\sum_{\theta\in\Theta^{\mathcal{R}}_{k,\nu}(p)}\mathrm{Val}(\theta) \\
&+ g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=2}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)g^{[n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon) \\
&\qquad \sum_{p=1}^{n-1}\sum_{q=0}^{p-1}M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\sum_{k=1}^{\infty}\sum_{\theta\in\Theta^{\mathcal{R}}_{k,\nu}(p)}\mathrm{Val}(\theta),
\end{aligned}
\tag{A5.3}
$$

which, by the definitions of $\mathbf{h}$, can be written as

$$
\begin{aligned}
g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} &= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\Big[\sum_{n=0}^{\infty}\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)\mathbf{h}_{n,\boldsymbol{\nu}} \\
&+\sum_{n=1}^{\infty}\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\sum_{p=n}^{\infty}\sum_{q=0}^{n-1}M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\mathbf{h}_{p,\boldsymbol{\nu}}+\sum_{n=2}^{\infty}\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\sum_{p=1}^{n-1}\sum_{q=0}^{p-1}M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\mathbf{h}_{p,\boldsymbol{\nu}}\Big].
\end{aligned}
\tag{A5.4}
$$

The terms in the second line of (A5.4) can be written as

$$\begin{aligned}
&\sum_{p=1}^{\infty}\Big(\sum_{q=0}^{p-1}\sum_{n=q+1}^{p} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)+\sum_{q=0}^{p-1}\sum_{n=p+1}^{\infty} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\Big)\mathbf{h}_{p,\boldsymbol{\nu}} \\
&\qquad = \sum_{p=1}^{\infty}\sum_{q=0}^{p-1} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\sum_{n=q+1}^{\infty}\Psi_n(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\,\mathbf{h}_{p,\boldsymbol{\nu}}
\end{aligned} \tag{A5.5}$$

and, by changing $p\to n$ and $n\to s$, we obtain

$$\begin{aligned}
&\sum_{n=1}^{\infty}\Big(\sum_{q=0}^{n-1} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\chi_0(\Delta^{[0]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\ldots\chi_q(\Delta^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\cdot \\
&\qquad\qquad \cdot\sum_{s=q+1}^{\infty}\chi_{q+1}(\Delta^{[q+1]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\ldots\psi_s(\Delta^{[s]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\Big)\mathbf{h}_{n,\boldsymbol{\nu}} \\
&\quad = \sum_{n=1}^{\infty}\sum_{q=0}^{n-1} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\chi_0(\Delta^{[0]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\ldots\chi_q(\Delta^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon))\,\mathbf{h}_{n,\boldsymbol{\nu}},
\end{aligned} \tag{A5.6}$$

where the identity (A5.1) has been used in the last line (with the correct interpretation of the term with $s=j+1$ explained after (A5.1)). By the definition of the matrices $\mathcal{M}^{[\le n]}(x;\varepsilon)$ one has

$$\sum_{q=0}^{n-1} M^{[q]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\chi_0(\Delta^{[0]}(x;\varepsilon))\ldots\chi_q(\Delta^{[q]}(x;\varepsilon))=\mathcal{M}^{[\le n]}(x;\varepsilon), \tag{A5.7}$$

so that, by inserting (A5.6) in (A5.3), after having used (A5.7), we obtain

$$\begin{aligned}
g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\left[\varepsilon\partial_{\boldsymbol{\varphi}}f(\boldsymbol{\varphi})\right]_{\boldsymbol{\nu}} &= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}\Big[\Big((\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2-\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big)+\mathcal{M}^{[\le n]}(\boldsymbol{\omega}\cdot\boldsymbol{\nu};\varepsilon)\Big]\mathbf{h}_{n,\boldsymbol{\nu}} \\
&= g(\boldsymbol{\omega}\cdot\boldsymbol{\nu})\sum_{n=0}^{\infty}(\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2\mathbf{h}_{n,\boldsymbol{\nu}}=\sum_{n=0}^{\infty}\mathbf{h}_{n,\boldsymbol{\nu}}=\mathbf{h}_{\boldsymbol{\nu}},
\end{aligned} \tag{A5.8}$$

so that the assertion is proved.

Remark. Note that at each step only absolutely converging series have been dealt with, so that the above analysis is rigorous and not only formal.

Acknowledgments. We are indebted to V. Mastropietro for many discussions and, in particular, to A. Giuliani for critical reading and several suggestions.

References

[B1] J. Bourgain, *Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE*, Internatational Mathematics Research Notices **11** (1994) 475–497.

[B2] J. Bourgain, *Construction of periodic solutions of nonlinear wave equations in higher dimension*, Geometric and Functional Analysis **5** (1995), 629–639.

[B3] J. Bourgain, *On Melnikov's persistency problem*, Mathematical Research Letters **4** (1997), 445–458.

[B4] J. Bourgain, *Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations*, Annals of Mathematics **148** (1998), no. 2, 363–439.

[Ba] J. C. A. Barata, *On formal quasi-periodic solutions of the Schrödinger equation for a two-level system with a Hamiltonian depending quasi-periodically on time*, Reviews in Mathematical Physics **12** (2000), no. 1, 25–64.

[BaG] M.V. Bartuccelli, G. Gentile, *Lindstedt series for perturbations of isochronous systems. A review of the general theory*, Reviews in Mathematical Physics **14** (2002), no. 2, 121–171.

[BGGM] F. Bonetto, G. Gallavotti, G. Gentile, Mastropietro, *Lindstedt series, ultraviolet divergences and Moser's theorem*, Annali della Scuola Normale Superiore di Pisa Classe di Scienze **26** (1998), no. 3, 545–593.

[C] C.-Q. Cheng, *Lower-dimensional invariant tori in the regions of instability for nearly integrable Hamiltonian systems*, Communications in Mathematical Physics **203** (1999), no. 2, 385–419.

[ChW] C.-Q. Cheng, S. Wang, *The surviving of lower-dimensional tori from a resonant torus of Hamiltonian systems*, Journal of Differential Equations **155** (1999), no. 2, 311–326.

[CrW] W. Craig and C.E. Wayne, *Newton's method and periodic solutions of nonlinear wave equations*, Communications in Pure and Applied Mathematics **46** (1993), no. 11, 1409–1501.

[E1] L.H. Eliasson, *Perturbations of stable invariant tori for Hamiltonian systems*, Annali della Scuola Normale Superiore di Pisa Classe di Scienze **15** (1988), no. 1, 115–147.

[E2] L.H. Eliasson, *Absolutely convergent series expansions for quasi-periodic motions*, Mathematical Physics Electronic Journal **2** (1996), paper 4, 1–33 (Preprint 1988).

[Fe] C. Fefferman, *Pointwise convergence of Fourier series*, Annals of Mathematics **98**, 551–571, 1973.

[G3] G. Gallavotti, *Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable hamiltonian systems. A review*, Reviews in Mathematical Physics **6**, 343–411, 1994.

[Ga] G. Gallavotti, *Twistless KAM tori*, Communications in Mathematical Physics **164** (1994), no. 1, 145–156.

[Ga01] G. Gallavotti: *Renormalization group in Statistical Mechanics and Mechanics: gauge symmetries and vanishing beta functions*, Physics Reports **352**, 251–272, 2001.

[Ga02] G. Gallavotti: *Exact Renormalization Group*, Paris IHP, 12 october 2002, Seminaire Poincaré, Editors B. Duplantier, V. Rivasseau, Institut H. Poincaré-CNRS-CEA.

[GG] G. Gallavotti, G. Gentile, *Hyperbolic low-dimensional tori and summations of divergent series*, Communications in Mathematical Physics **227** (2002), no. 3, 421–460.

[GBG] G. Gallavotti, F. Bonetto, G. Gentile, *Aspects of the ergodic, qualitative and statistical properties of motion* Springer–Verlag, Berlin, 2004.

[Ge] G. Gentile, *Quasi-periodic solutions for two-level systems*, Communications in Mathematical Physics **242** (2003), no. 1–2, 221–250.

[GM1] G. Gentile, V. Mastropietro, *Methods for the analysis of the Lindstedt series for KAM tori and renormalizability in classical mechanics. A review with some applications*, Reviews in Mathematical Physics **8** (1996), no. 3, 393–444.

[GM2] G. Gentile, V. Mastropietro, *Construction of periodic solutions of nonlinear wave equations with Dirichlet boundary conditions by the Lindstedt series method*, to appear on Journal de Mathématiques Pures et Appliquées.

[GMP] G. Gentile, V. Mastropietro, M. Procesi, *Periodic solutions for completely resonant wave equations*, Preprint, 2004.

[JLZ] A. Jorba, R. de la Llave, M. Zou, *Lindstedt series for lower-dimensional tori*, in *Hamiltonian systems with three or more degrees of freedom* (S'Agaró, 1995), 151–167, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 533, Ed. C. Simó, Kluwer Acad. Publ., Dordrecht, 1999.

[Ka] T. Kato, *Perturbation theory for linear operators*, Grundlehren der Mathematischen Wissenschaften, Band 132, Springer-Verlag, Berlin-New York, 1976.

[Ku] S.B. Kuksin, *Nearly integrable infinite-dimensional Hamiltonian systems*, Lecture Notes in Mathematics 1556, Springer-Verlag, Berlin, 1993.

[KP] S.B. Kuksin, J. Pöschel, *Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation*, Annals of Mathematics **143** (1996), no. 1, 149–179.

[LW] R. de la Llave, C.E. Wayne, *Whiskered and low dimensional tori in nearli integrable Hamiltonian systems*, Mathematical Physics Electronic Journal, 2004.

[Me1] V.K. Mel'nikov, *On certain cases of conservation of conditionally periodic motions under a small change of the Hamiltonian function*, Doklady Akademii Nauk SSSR **165** (1965), 1245–1248; english translation in Soviet Mathematics Doklady **6** (1965), 1592–1596.

[Me2] V.K. Mel'nikov, *A certain family of conditionally periodic solutions of a Hamiltonian systems*, Doklady Akademii Nauk SSSR **181** (1968), 546–549; english translation in Soviet Mathematics Doklady **9** (1968), 882-886.

[Mo] J. Moser, *Convergent series expansions for quasi-periodic motions*, Mathematische Annalen **169** (1967), 136–176.

[P1] J. Pöschel, *On elliptic lower-dimensional tori in Hamiltonian systems*, Mathematische Zeitschrift **202** (1989), no. 4, 559–608.

[P2] J. Pöschel, *Quasi-periodic solutions for a nonlinear wave equation*, Commentarii Mathematici Helvetici **71** (1996), no. 2, 269–296.

[RS] M. Reed, B. Simon, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press, New York-London, 1978.

[R] H. Rüssmann, *Invariant tori in non-degenerate nearly integrable Hamiltonian systems*, Regular and Chaotic Dynamics **6** (2001), 119–204.

[T] D.V. Treshchëv, *A mechanism for the destruction of resonance tori in Hamiltonian systems*, Rossiĭskaya Akademiya Nauk. MatematicheskiĭSbornik **180** (1989), no. 10, 1325–1346; english translation in Mathematics of the USSR-Sbornik **68** (1991), no. 1, 181–203.

[WC] S. Wang, C.-Q. Cheng, *Lower-dimensional tori for generic Hamiltonian systems*, Chinese Science Bulletin **44** (1999), no. 13, 1187–1191.

[Wa] C.E. Wayne, *Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory*, Communications in Mathematical Physics **127** (1990), no. 3, 479–528.

[XY] J. Xu, J. You: *Persistence of lower dimensional tori under the first Melnikov's non-resonance condition*, Journal de Mathématiques Pures et Appliquées **80** (2001), no. 10, 1045–1067.

[Y] J. You: *Perturbations of lower-dimensional tori for Hamiltonian systems*, Journal of Differential Equations **152** (1999), no. 1, 1–29.